

Portland State University

PDXScholar

Dissertations and Theses

Dissertations and Theses

1-1-1987

Segment Congruence Analysis: An Information Theoretic Approach

Jamshid Hosseini-Chaleshtari
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Let us know how access to this document benefits you.

Recommended Citation

Hosseini-Chaleshtari, Jamshid, "Segment Congruence Analysis: An Information Theoretic Approach" (1987). *Dissertations and Theses*. Paper 797.
<https://doi.org/10.15760/etd.797>

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

SEGMENT CONGRUENCE ANALYSIS:
AN INFORMATION THEORETIC APPROACH

by
JAMSHID HOSSEINI-CHALESHTARI

A dissertation submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY
in
SYSTEMS SCIENCE

Portland State University

© 1987

TO THE OFFICE OF GRADUATE STUDIES AND RESEARCH:

The members of the Committee approve the dissertation
of Jamshid Hosseini-Chaleshtari presented June 29, 1987.


Robert R. Harmon, Chairman


Martin Zwick


Thomas R. Gillpatrick


Alan R. Raedels


John S. Oh
Office of Graduate Studies Representative

APPROVED: 

Martin Zwick, Coordinator, Systems Science


Bernard Ross, Vice Provost for Graduate Studies

To my father, Khodayar, and my mother, Tooba,
whose unconditional love has been a
guiding force in my life and an
inspiration to me as a parent.

To my lovely wife, Pamela, whose patience and
loving support provided the peace of
mind I needed to undertake this project.

And to my children, Nema and Arman, who give
meaning and purpose to my life.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to the many individuals who provided assistance and encouragement to me throughout my studies. In particular, I am immeasurably indebted to Dr. Robert Harmon whose support I could rely on and whose guidance kept me on track. He was always willing to listen and offer well placed and practical advice. I have been most fortunate to know him as a teacher, a friend, and a colleague.

Words cannot express my respect and gratitude for Dr. Martin Zwick whose quest for excellence and perfection has had a profound effect on my academic and personal life. He introduced me to the general systems methodology area and, during the past several years, he has helped me to develop scientific and critical thinking. I am proud and honored to have had the benefit of knowing and working with such a true scholar.

My grateful appreciation also extends to the other members of my dissertation committee, Drs. Thomas Gillpatrick, Alan Raedels, and John Oh, for their generous help and kindly advice. Their discussions and comments

regarding the nature and direction of this work were both stimulating and constructive.

Other colleagues and friends also assisted me in many ways. I am especially thankful to my friend Muhannad Khawaja who was an excellent listener and facilitated the acquisition of the data I used for this study. To that end, I am thankful to Dr. Gill Peach and Ms. Karen Schoch of Pacific Power and Light Company and Mike Austin and Erick Westman of Bonneville Power Administration who were instrumental in the process.

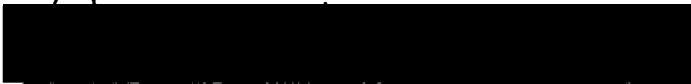
I am also indebted to Mr. Wess Brenner, Ms. An Hotan, and Mr. Dave Sato of the Academic Computing at Portland State University for their technical support and good humor. Ms. Dawn Kuenle of the System Science Department, Ms. Maureen Eldred and Berni Pilip of the Office of Graduate Studies and, Ms. Sally Lopez all helped facilitate the operational aspects of my dissertation for which I am grateful.

Finally, my beloved family, my lovely wife, Pamela, my dearest sons Nema and Arman, my brother Behzad and my sister Maryam, deserve immeasurable gratitude for keeping me sane throughout this work. Not many people are fortunate enough to have so many individuals on their side and so much encouragement as I had.

AN ABSTRACT OF THE DISSERTATION OF Jamshid Hosseini-
Chaleshtari for the Doctor of Philosophy in Systems Science
presented June 29, 1987.

Title: Segment Congruence Analysis: An Information
Theoretic Approach

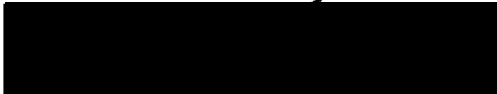
APPROVED BY MEMBERS OF THE DISSERTATION COMMITTEE:


Robert R. Harmon, Chairman


Martin Zwick


Thomas R. Gillpatrick


Alan R. Raedels


John S. Oh,
Office of Graduate Studies Representative

During the past three decades, marketers have developed various methods to partition the total market into subgroups of consumers whose responses to certain marketing strategies are homogeneous within the subgroups and heterogeneous across them. Usually, this analysis begins by identifying a segmentation variable (a priori and/or by

clustering) which characterizes consumer behavior. A particular class of this variable is identified as the target segment for which a marketing strategy is to be developed. This target segment is then analyzed in terms of descriptor variables.

When there are several possible segmentation variables, marketers must investigate the ramifications of their potential interactions. These include their mutual association, the identification of the best (the distinguished) segmentation variable and its predictability by a set of descriptor variables, and the structure of the multivariate system(s) obtained from the segmentation and descriptor variables. This procedure has been defined as segment congruence analysis (SCA) (Green and Carmone 1977).

Traditionally, this has been done using general log-linear and logit models. This study utilizes the information theoretic approach, as well as the log-linear/logit approach, to address a variety of research questions in segment congruence analysis. It is shown that the information theoretic approach expands the scope of SCA and offers some advantages over traditional methods.

Data obtained from a survey conducted by the Bonneville Power Administration and Northwest utilities (PNWRES, the 1983 Pacific Northwest Residential Survey) is used to demonstrate the efficacy of the information theoretic and the log-linear/logit approaches and compare

these two methods. The survey was designed to obtain information on energy consumption habits, attitudes toward selected energy issues, and the conservation measures utilized by the residents in the Pacific Northwest

The analyses are performed in two distinct phases. Phase I includes assessment of mutual association among segmentation variables and four methods (based on different information theoretic functions) for identifying candidates for the distinguished variable. Phase II addresses the selection and analysis of the distinguished variable. This variable is selected either a priori or by assessment of its predictability from (segmentation or exogenous) descriptor variables. The relations between the distinguished variable and the descriptor variables are further analyzed by examining the predictability issue in greater detail and by evaluating structural models of the multivariate systems.

SPSS^X (1986) is used to 1) combine clusters of conceptually related variables in the original survey data into aggregate variables; 2) perform general log-linear, hierarchical log-linear, and logit analyses; and 3) produce input files for the information theory program (a FORTRAN 77 program).

The methodological conclusions of this study are that the information theoretic and log-linear methods have deep similarities. Therefore, it is possible to utilize the readily available information theoretic measures in

developing log-linear models. In this study, several new log-linear models, which have not been used in the past, are developed based on their information theoretic counterparts.

PNWRES analyses produced intuitively plausible results. In Phase I, energy related awareness, behavior, perceptions, attitudes, and electricity consumption were identified as candidate segmentation variables. In Phase II, using exogenous descriptor variables, electricity consumption was selected as the distinguished variable. The analysis of this variable indicated that the demographic factors, type of dwelling, and geoclimatic environment are among the most important determinants of electricity consumption.

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGEMENTS.....	iv
TABLE OF CONTENTS.....	vi
LIST OF TABLES.....	x
LIST OF FIGURES.....	xv
CHAPTER	
I INTRODUCTION.....	1
Objectives of the Study.....	4
Motivation for the Study.....	11
II REVIEW OF THE LITERATURE.....	14
An Overview of Market Segmentation.....	15
Segment Congruence Analysis.....	19
The Role of Contingency Table Analysis	
The Log-Linear/Logit Approach.....	28
Relevant Applications of the General Log-Linear Model	
Relevant Applications of the Logit Model	

		vii
CHAPTER		PAGE
	The Information Theoretic Approach.....	37
	Some Important Properties of Information Theory Information Theoretic Definitions of a Leading Part Statistical Significance of Informational Measures Information Theory as a Tool for Segment Congruence Analysis	
	The General Probability and Chi-Square Approach.....	50
	Summary.....	53
III	RESEARCH QUESTIONS, METHODOLOGY, AND PROCEDURES.....	56
	Research Questions.....	56
	Phase I: Analysis of the Overlap Phase II: Identification and Analysis of the Distinguished Variable	
	Methodology.....	72
	Information Theoretic Methodology Computer Program The General Log-Linear/Logit Approach Summary of the Methodology	
	The Data and Procedures.....	111
	Description of the Study Selection and Aggregation of Variables	

		viii
CHAPTER		PAGE
IV	RESULTS.....	123
	Analysis of the Segmentation	
	Variables.....	126
	Existence and Degree of Mutual Association Among Segmentation Variables	
	Identification of Candidates for the Distinguished Variables	
	Determination and Analysis of the	
	Distinguished Segmentation Variable	
	Using the Other Segmentation	
	Variables as the Descriptor	
	Variables.....	143
	Electricity Consumption as the Distinguished Segmentation Variable (A Priori)	
	Analytical Identification of the Distinguished Segmentation Variable	
	Determination and Analysis of the	
	Distinguished Segmentation Variable	
	Using Exogenous Descriptor	
	Variables.....	165
	Summary of Analysis.....	176
V	CONCLUSIONS AND FUTURE DIRECTIONS.....	181
	Findings of the Study.....	181

		ix
CHAPTER		PAGE
	Contributions of the Study.....	189
	Contribution to Segment Congruence Analysis Contribution to Market Segmentation Contribution to Log-Linear Modeling Contribution to Information Theoretic Modeling	
	Limitations.....	196
	Directions for Further Research.....	198
BIBLIOGRAPHY.....		200
APPENDIX I	General Log-Linear Models.....	213
APPENDIX II	Entropy and Information Theoretic Concepts.....	220
APPENDIX III	Pacific Northwest Residential Consumption Survey (Spring 1983) - Code Book.....	229
APPENDIX IV	Pacific Northwest Residential Consumption Survey (1983) - List of Variables.....	286
APPENDIX V	Sample Input/Output for SYSENT, The Information Theoretic Data Analysis Package.....	299

PLEASE NOTE:

**Page(s) not included with original material
and unavailable from author or university.
Filmed as received.**

U·M·I

TABLE

PAGE

VIII	Information Theoretic and Log-Linear Mathematical Expressions Used for Segment Congruence Analysis (Phase II).....	109
IX	Comparison of the Two Approaches to Segment Congruence Analysis.....	110
X	List of Variables Prior to Aggregations...	120
XI	List of Variables (or Clusters) After Aggregations.....	122
XII	List of the Segmentation Variables.....	127
XIII	Assessment of the Mutual Association.....	128
XIV	Assessment of the Contribution of Each Variable to the Mutual Association.....	131
XV	Assessment of the Contribution of Each Variable to the Joint Variability in the System of Segmentation Variables...	136
XVI	Assessment of the Contribution of Each Variable to the Total Uncertainty in the System of Segmentation Variables...	139
XVII	Assessment of the Number of Significant Dyadic Relations.....	142
XVIII	Predictability of a Particular Class vs. All Other Classes of the Distinguished Variable.....	146

		xii
TABLE		PAGE
XIX	Prioritized Effect of the Other Segmentation Variables on Energy Consumption	149
XX	Dependency Analysis of the Four-Variable System of GENATT, ENRATT, PERCEPT, and KWHUSE.....	151
XXI	Dependency Analysis of the Six-Variable System of Segmentation Variables.....	155
XXII	Predictability of Candidates for the Distinguished Segmentation Variable....	158
XXIII	Predictability of a Particular Class vs. All Other Classes of Energy Related Behavior.....	159
XXIV	Prioritization of the Effect of the Other Segmentation Variables on Energy Related Behavior (BEHAVE).....	161
XXV	Dependency Analysis of the Four-Variable System of BEHAVE, AWARE, KWHUSE, and ENRATT.....	163
XXVI	List of the Exogenous Descriptor Variables.....	166
XXVII	Predictability of Candidates for the Distinguished Segmentation Variable Using Exogenous Descriptor Variables...	167

TABLE

PAGE

XXVIII	Predictability of a Particular Class vs. All Other Classes of the Distinguished Variable, (KWHUSE).....	169
XXIX	Prioritized Effect of the Exogenous Descriptor Variables on Energy Consumption.....	171
XXX	Dependency Analysis of the Four-Variable System of KWHUSE, DEMOG, TYPDWEL, and CLIMGEO.....	172
XXXI	Dependency Analysis of the Distinguished Variable and the Exogenous Descriptor Variables.....	174
XXXII	Analysis of the Mutual Association Among Segmentation Variables and Identification of Candidates for the Distinguished Segmentation Variable....	177
XXXIII	Analysis Using Other Segmentation Variables as Descriptor Variables (Electricity Consumption as the Distinguished Variable (<u>a priori</u>)).....	178
XXXIV	Analysis Using Other Segmentation Variables as Descriptor Variables (Analytical Identification and Analysis of the Distinguishd Variable).....	179

TABLE

PAGE

XXXV	Analysis Using Exogenous Descriptor Variables.....	180
XXXVI	Significance of the Results of Segment Congruence Analysis to the PNWRES Study: Phase I.....	184
XXXVII	Significance of the Results of Segment Congruence Analysis to the PNWRES Study (Phase II: <u>A Priori</u> Distinguished Variable).....	185
XXXVIII	Significance of the Results of Segment Congruence Analysis to the PNWRES Study (Phase II: Other Segmentation Variables as Descriptor Variables (Analytical))..	186
XXXIX	Significance of the Results of Segment Congruence Analysis to the PNWRES Study (Phase II: Exogenous Descriptor Variables).....	187
XXXX	Actual and Expected Cell Frequencies for a 2X2 Contingency Table.....	215

LIST OF FIGURES

FIGURE	PAGE
1. A Schematic View of the Total Study.....	20
2. Segment Congruence Analysis Process as Proposed by Green and Carmone (1977)...	22
3. Spatial Representation of Transmission.....	74
4. A System of Four Variables.....	76
5. Systematic Entropy of a Four-Variable System.....	78
6. Determination of Variable A With the Knowledge of the Descriptor Variables X, Y, and Z.....	84
7. Prioritization of the Contributions of Variables X, Y, and Z to Determine Variable A.....	88
8. A Sample Dependency Analysis Diagram.....	92
9. Flow Chart for Selection and Aggregation of Research Variables.....	117
10. Flow Chart of the Log-Linear/Logit Analysis.....	118
11. Flow Chart of the Information Theoretic Analysis.....	119

		xvi
FIGURE		PAGE
12.	Condensed Flow Chart of Data Analysis.....	125
13.	Graphical Representation of the Dependency Analysis for the Four-Variable System of GENATT, ENRATT, PERCEPT, and KWHUSE.....	153
14.	Graphical Representation of the Dependency Structures for the Six-Variable System of Segmentation Variables.....	156
15.	Graphical Representation for the Dependency Structures for the Four- Variable System of BEHAVE, AWARE, KWHUSE, and ENRATT.....	164
16.	Graphical Representation of the Dependency Structures for the Four-Variable System of KWHUSE, DEMOG, TYPDWEL, and CLIMGEO.....	173
17.	Graphical Representation of the Dependency Structures for the System of Distinguished and Exogenous Variables.....	175

CHAPTER I

INTRODUCTION

Market segmentation is defined as the grouping of potential customers into subgroups such that their responses to certain market related stimuli (i.e., price, promotion, product, and distribution strategies) are homogeneous within each group and heterogeneous between groups (Lillien and Kotler, 1983). Market segmentation has been an extremely popular subject since late 1940's and early 1950's, and gained prominence through pioneering work by Wendell Smith (1956).

The process of market segmentation begins by identifying a variable based on which the market is to be partitioned. Then, different classes of this variable are selected as different market segments. Next, these segments are examined for homogeneity within and heterogeneity across them (e.g., by analyzing variance within and between segments), and stability (e.g., by utilizing the multinomial fixed or variable Markov models) (Wind, 1978). Then, the predictability of each segment by a set of descriptor variables is assessed (e.g., through multiple regression

analysis) (Blattberg and Sen, 1976), or its discrimination from other segments is examined (e.g., using multiple discriminant analysis) (Perreault, et al 1979).

In certain market segmentation studies, several variables, instead of one, may be deemed suitable for partitioning the market. In such cases, the primary concern involves the handling of this multivariate problem-- e.g., the identification of a variable to serve as the distinguished segmentation variable, the predictability of this variable given the knowledge of the values, or classes, of other variables, etc.

"Traditionally, researchers have selected some variable or a battery of variables and proceeded from there." (Green 1977, p. 66) This a priori approach either reflects the manager's experience or hunch, or implies the existence of a highly developed body of theory or past research. One disadvantage of this approach is that one's prior convictions can act more as a set of blinders than as guide to further study and understanding. In addition, "regardless of the complexity of reality, human beings find it difficult to classify objects by more than two or three characteristics at a time. If reality requires greater complexity, we are severely constrained by our own conceptual limitations." (Frank et al 1972, p. 122).

Strategy considerations should facilitate the choice of segmentation variables. However, even based on strategic

issues cases can arise in which several candidate bases are plausible choices. What is needed, then, is a method that will:

(1) show how closely related various clusters obtained from alternative bases are, (2) test whether some clusters are independent of the remaining ones, and (3) find the clustering that exhibits the highest contribution to the mutual association among the subset of clusters that evince significant associations in the first place. Furthermore, even if the researcher has settled in advance on a distinguished set of variables to serve as the segmentation base, s/he may be interested in finding out how well the clustering obtained from this distinguished set can be predicted by some function derived from the other clusters (Green, 1977, p. 67)

Analysis of interrelationships among multivariate segmentation problems has been defined as segment congruence analysis (Green and Carmone 1977). The basic premise of this method is that, if mutual association exists among segmentation bases, knowledge of a population unit's membership in one segment may increase or decrease the likelihood of his/her membership in another segment.

Segment congruence analysis is useful to the researcher in a number of ways: 1) It enables the elimination of unnecessary segmentation variables. 2) It helps reduce the overlap in marketing mix strategies directed to different segments. 3) It is instrumental in the identification of the "best" (or the "distinguished") segmentation variable.

OBJECTIVES OF THE STUDY

Not unlike other multivariate problems, segment congruence analysis entails a great deal of complexity. This is due to the fact that, here, several variables are considered simultaneously rather than individually or sequentially. This study investigates two methods of analyzing multivariate problems in the context of segment congruence analysis; the information theoretic approach (i.e., the proposed method) and the log-linear/logit approach (i.e., the traditional method).

The main objectives of this study include: (1) enrichment of segment congruence analysis by introducing numerous relevant questions which have not previously been addressed, (2) introduction and application of information theory as a method to address all of the relevant questions in this area, and (3) comparison of this proposed method with the log-linear/logit models as traditionally used for the purpose of segment congruence analysis.

Data collected through a survey conducted by the Bonneville Power Administration (BPA) in collaboration with the local utilities in the Northwest region (i.e., the Pacific Northwest Residential Energy Consumption Survey, PNWRES) provided the basis for an empirical application of segment congruence analysis. This data was used to

demonstrate the usefulness of segment congruence analysis (including the newly introduced research questions), the efficacy of the information theoretic approach, and its comparison with the log-linear/logit approach.

The PNWRES study was designed to gain better understanding of energy use habits, attitudes, and conservation activities in the Pacific Northwest region. The BPA decision makers are concerned with energy related issues such as the level of energy usage (i.e., for their energy planning activities) and promotion of conservational programs (particularly, during a period of energy shortage). For both of these purposes, the knowledge of consumers' demographic, socioeconomic, attitudinal, behavioral, and other relevant dimensions, as well as the characteristics of their residence are essential. It is through this knowledge that the energy usage can be predicted, or the success of a conservation oriented program can be estimated prior to its actual implementation.

PNWRES data was first utilized to determine a set of attitudinal, behavioral, perceptual, awareness, and consumption variables as potential segmentation variables. The demographic and socioeconomic profiles of the residents (e.g., income, education, ethnicity, etc.) as well as the characteristics of residential units (e.g, geographic, climatic, size, level of insulation, etc.) were identified as possible descriptor variables.

Each variable was obtained by grouping a number of related variables, in the original data, into a single variable (i.e., using a cluster analysis procedure). In this context, clustering means grouping of similar classes of several variables into corresponding classes of a combined variable. Table I contains a listing of these (clustered) variables.

Segment congruence analysis as proposed by Green and Carmone (1977) and expanded here, includes two phases: (1) analysis of the multivariate system of segmentation variables (i.e., analysis of mutual association and identification of candidates variables for the distinguished segmentation variable), and (2) identification and analysis of the best segmentation variable among the candidate distinguished variables (i.e., predictability and identification of the best descriptor variables).

TABLE I
A CONDENSED LIST OF VARIABLES USED IN THIS
STUDY

Type of Variable	Abbreviated Name
<u>Segmentation Variables</u>	
General attitudes-- economy and environment	(GENATT)
Energy related attitudes	(ENRATT)
Energy related behavior	(BEHAVE)
Energy (comfort/discomfort) perceptions	(PERCEPT)
Energy-saving-development funding awareness	(AWARE)
Energy usage in kilo-watt-hours	(KWHUSE)
<u>Descriptor Variables</u>	
Geographic/climatic environment	(CLIMGEO)
Type of dwelling	(TYPDWEL)
Status of the resident in terms of renter/owner	(RENTOWN)
Demographic cluster	(DEMOG)
Size of the dwelling	(DWELSIZE)
Level of insulation of the residence	(INSUL)

In the context of energy-usage-prediction or conservation-program-evaluation some or all of the following questions are addressed when analyzing PNWRES data:

- a) Do the attitudinal, perceptual, behavioral, awareness, and consumption variables show overlap (i.e., is it possible to gain knowledge about the class of one variable if the classes of other variables are known)?
- b) Given that the overlap exists among these variables, which one variable should be considered as a candidate distinguished variable (e.g., which variable contribute the most toward the overlap etc.)?
- c) Given a number of candidate distinguished segmentation variables, which one is best determined with the knowledge of the descriptor variables)?
- d) Given a best segmentation variable, how well are different classes of this variable predicted by the descriptor variables?
- e) Can the number of descriptor variables be reduced without significant loss in the predictive power of the prediction model?
- f) What is the structure of the multivariate system composed of the descriptor variables and segmentation variable (i.e., can a simpler model of the system be developed?)

Decision makers at BPA, as well as local utilities in the region and even the energy industry as a whole, can benefit from such analyses in several critical ways. First, by analyzing the overlap, they will know whether the knowledge of certain variables, which can be measured with higher validity and reliability, can contribute to the knowledge of other, more vague, variables. For instance,

consider a situation in which the consumers attitudes toward energy use are sought for a conservation program. If it is established that knowledge of energy consumption contributes significantly to the knowledge of the consumers' energy use attitudes, it may be advisable to measure the more objective energy consumption variable and assess the attitudinal variable based on the energy consumption variable.

Second, the issue of predictability of segmentation variables is essential in market segmentation studies. For instance, assuming that the result of the preliminary segment congruence analysis reveals that general attitudes about the environment, energy related behavior, and energy consumption are the candidates for distinguished segmentation variables. Predictability of each of these candidate variables is then assessed and the variable that is best predicted is designated as the best (analytically derived) best segmentation variable.

Third, once a set of descriptor variables are identified, particularly if a large number of such variables are considered, it is necessary to eliminate variables which do not contribute significantly to the prediction of the segmentation variable. Unnecessary variables are not only costly in terms time and computational aspects, but cloud the analysis, and in some cases lead to incorrect analyses and conclusions.

In general, multivariate problems, including simultaneous assessment of variables in market segmentation studies, are very complex. Segment congruence analysis provides a framework to deal with this complexity which is intuitively plausible. Green and Carmone (1977) used the log-linear and logit approach to segment congruence analysis. The present undertaking employs information theory to further develop this framework.

The proposed information theoretic methodology for segment congruence analysis enables the researcher in the field of energy marketing to gain increased understanding about different segments within the residential consumers and their accessibility. Segment congruence analysis is obviously not limited to energy related issues. Any industry in which a dominant segmentation variable is not readily available can use segment congruence analysis. Even when a distinguished segmentation variable exists, these analyses will enhance understanding of the market segments and their related descriptor variables.

In the present study, several procedures available in the SPSS^X (1986) package are used to 1) combine classes of conceptually related variables into aggregate classes and produce variables with clustered categories; 2) perform general log-linear, hierarchical log-linear, and logit analysis in addressing the above issues; and 3) produce input files for the information theory program from the raw

data file. A FORTRAN 77 computer program is developed to perform the information theoretic analysis.

MOTIVATION FOR THE STUDY

As an area of inquiry, market segmentation is one of the most widely investigated areas in marketing profession. It is the foundation of the modern marketing concept in which consumer needs are among the primary considerations in developing marketing strategy. Market segmentation has enjoyed considerable attention in recent years due to, at least, the following reasons:

1. As a way of thinking about underlying strategy of a corporation's marketing program, the concept can be pervasive -- it can affect every component of the program.
2. The concept of segmentation is unusual in that it is perceived both as an important building block in the marketing program of many firms and as a stimulus to the intellectual curiosity of research professionals in both the business and academic communities (Frank et al 1972, p. 246).

Segmentation studies using multiple variables are becoming more common as the marketing practitioners realize the need for further refining of market segments. When multiple variables are used, an individual may belong to two or more of the segments identified. For instance, the individual may seek fresh breath as well as decay prevention in choosing a toothpaste. Looking at this issue from the view point of product positioning, a specific brand of

chewing gum may compete with other brands of gum as a candy substitute as well as competing with certain brands of mouthwash or with different brands of toothpaste or even cigarettes! (Arabie et al 1981)

In both segmentation and positioning contexts, marketing theoreticians and practitioners must deal with overlapping market segments. Failure to recognize that an individual belonging to the "socially aware" segment may also belong to the "health oriented" segment in the toothpaste market segmentation can result in the wrong choice in developing the marketing mix strategy. This problem becomes particularly critical when the overlap between the two segments is large. The same can be said with respect to product positioning and mixed (segmentation and positioning) situations.

Segment congruence analysis, therefore, is extremely important in situations where multiple segmentation variables are used. These situations are among the most difficult problems to deal with due to the level of complexity involved.

The proposed information theoretic method is valuable because there exists no other single method (or family of techniques) that can effectively deal with all of the aspects of segment congruence analysis. The following chapter reviews the current literature in segment congruence analysis and its relevant topics. It will be shown that,

not only there is no unified family of techniques that can be used to address these issues, but some of the techniques are based on assumptions which may not be suitable for categorical data. Categorical data are the most widely used types of data available for defining different market segments.

Watanabe (1969) refers to information theory as a powerful tool for analyzing the organizational structure. This is because information theory offers a methodology for quantifying organization or patterning (Miller, 1953). "This creates great fascination for information theory as applied to psychology [and related social sciences]." (Attneave, 1959) This dissertation provides some additional support for these assertions.

CHAPTER II

REVIEW OF THE LITERATURE

The process of market segmentation begins with the identification of a variable (e.g., deal proneness) or a set of variables (e.g., deal proneness and brand loyalty) as the segmentation set. These variables are drawn from a segmentation base (or basis)¹. The identification of segmentation set of variables (or base) has traditionally been done either on a priori or a posteriori basis, or a hybrid of the two. (These approaches will be discussed shortly.)

Then, certain classes of a single variable (e.g., the deal prone or the non-deal prone) or combinations of classes

¹ The definition of a base, and its distinction from a group of variables (or even one variable), is context dependent. In general, a segmentation base is comprised of such factors as demographic/ socioeconomic characteristics, consumption patterns, personality traits, and attitudes, perceptions, and preferences. Frank, et al (1972 pp.26-28) treat variables, such as stage of life cycle and social class, as bases. Furthermore, in discussing the criteria for evaluating alternative bases for segmentation, they treat variables and bases interchangeably. Both Green and Carmone (1977) and Van Auken and Lonial (1984) treat specific variables as bases, in segment congruence analysis.

of the set of variables (e.g., deal prone and brand loyal, deal prone and brand disloyal, etc.) are identified as the target groups (i.e., target segments). Next, a profile of each target segment is created on the basis of demographic, socioeconomic, geographic, etc., characteristics of its members. Subsequently, an attempt is made to reach (access) these segments through different marketing strategies (a market segment is called accessible if one can identify and/or target the members of that segment effectively).

AN OVERVIEW OF MARKET SEGMENTATION

Market segmentation is one of the most intensively investigated topics among both marketing researchers and practitioners today. A bibliography of publications in market segmentation during the past three decades would probably sum to several hundred pages. However, there has been relatively little substantial research to put in perspective, among certain other important issues, the role and importance of multivariate segmentation analyses.

One typology for market segmentation involves different approaches by which the segmentation variables are identified. The most common approaches include a priori, a posteriori (also referred to as post-hoc or clustering based), or hybrid (mixed) segmentation schemes (Wind 1978). In a priori segmentation, the researcher chooses some

cluster-defining descriptor in advance, such as usage rate of a product. Respondents are then classified into clusters (or segments), such as heavy-users and light-users, and are further examined regarding their differences in other characteristics, such as demographics or benefits sought.

a posteriori (i.e, post-hoc or clustering-based) segmentation, on the other hand, starts by clustering respondents according to the similarity of their multivariate profiles regarding such characteristics as purchase behavior, attitudes, personality, life styles, and/or any other meaningful set of variables. Following this, the segments are examined for differences in other characteristics, not used in the original profile definition.

Some studies require a hybrid of the two approaches. For instance, respondents could first be grouped based on their usage rate, a priori, and then a clustering procedure could be employed to see if other segments revealing common characteristics, such as different benefit-seeking groups, emerge.

Another typology for market segmentation concerns the methods of analysis for segmentation and/or descriptor variables. These include multivariable and multivariate analysis. (Frank, et al 1972) Multivariable analysis refers to situations where a single segmentation variable is considered as the criterion (dependent) variable and several

other variables are considered as descriptor (independent) variables. In this case, if multiple segmentation variables are identified, they are treated separately and independently of each other. (i.e., separate functions are developed for assessing the relationship between each segmentation variable and (a set of) descriptor variables.) Several quantitative techniques are available for this general conceptualization².

The second approach, known as multivariate, requires that several segmentation variables be considered simultaneously. This approach, in general, is more complex than the multivariable approach. Here, first, a number of segmentation variables are analyzed (simultaneously) and a distinguished variable is selected. Next, relevant descriptor variables, which best determine or predict the segmentation variable, are identified.

One group of techniques is primarily suitable for simultaneous analysis of segmentation problems with intervally scaled dependent variables³. A major shortcoming

2 For instance, regression analysis, multiple discriminant analysis, multiple classification analysis (MCA), analysis of variance (ANOVA), automatic interaction detection (AID), conjoint analysis, etc.

3 These problems have been addressed in the literature using, among others, canonical correlation analysis (Alpert 1971), multivariate analysis of variance (MANOVA), additive clustering (ADDCLUS) (Arabie et. al 1981.), mathematical programming clustering (MAPCLUS) (Arabie 1977), and nonparametric multidimensional scaling. (Green and Rao 1972).

of some of these and traditional techniques, such as regression analysis, automatic interaction detection (AID), ANOVA, and their extensions, is that most of the practical applications would violate one or more of their assumptions. Green, et al (1977) and DeSarbo and Hildebrand (1980), attribute this problem to:

1. The assumption of normality for the criterion variable is unreasonable.
2. The criterion variable does not display constant variance (homoscedasticity) across variations in the predictors.
3. The model's predictions could fall outside of the range of 0 to 1.

"Multidimensional scaling methods, [on the other hand,] offer no analytical solution, but rather act as a transformation function to provide spatial structures underlying a data base through a series of interactive procedures ..." (Best 1975, p.16-17) Therefore, these techniques cannot be employed, for instance, to predict the behavior (or response to marketing stimuli) of a particular segment.

Another group of techniques are intended for all discrete (nominal or ordinal) data for both dependent and independent variables. It is this group of techniques that is the subject of segment congruence analysis and is explored throughout this study. In the next section, segment congruence analysis is introduced and relevant prior

research in this area is explored. The second section describes the role of multivariate contingency tables in segment congruence analysis. Section three provides a more detailed discussion of the log-linear and logit approach to simultaneous contingency table analysis. Then, in section four, the information theoretic approach is introduced along with some of the more recent applications in this area. Section five provides a brief introduction to the general probability and chi-square analysis. Finally, a summary of this chapter is presented in section six. Figure 1 presents a schematic view of this study.

SEGMENT CONGRUENCE ANALYSIS

Segment congruence analysis, as proposed by Green and Carmone (1977), was based on techniques for analyzing categorical data. The basic procedures for segment congruence analysis are described by Green (1977) as follows (Here, a "cluster" refers to a class of a variable (X_i , $i=1$ or 2 or 3 ...), or a composite class of several variables expressed simultaneously ($X_i Y_j Z_k \dots$, $i=1$ or 2 or 3 ..., $j=1$ or 2 or 3, etc.); while, a "clustering" refers to a variable (X_i , $i=1,2,\dots$), or a set of variables ($X_i Y_j Z_k \dots$, $i=1,2,3,\dots$, $j=1,2,3,\dots$, etc.):

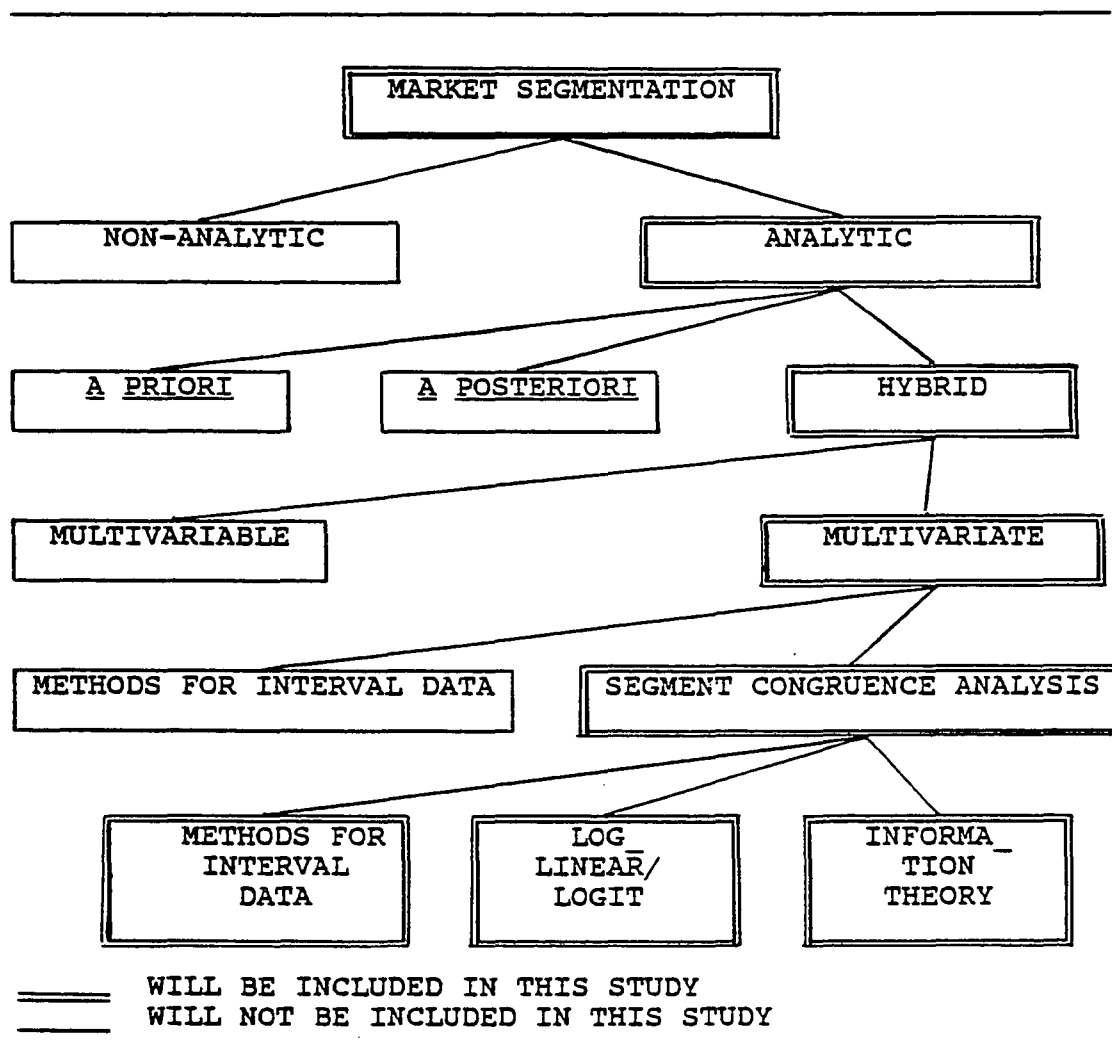


Figure 1. A schematic view of the total study.

1. Each candidate segmentation base (battery of variables) is independently used to cluster the respondents into some specified number of groups. The number of clusters so formed may or may not be the same across bases, depending upon the researcher's objectives.
2. Each clustering serves as a categorical variable (where the clusters represent the categories), and a multidimensional contingency table is formed.
3. A variety of tests can then be carried out on the multiway table to see which clusters are related to which other ones.
4. One clustering can be chosen as the distinguished clustering and a model can be developed for predicting this clustering from clusters obtained from other bases. In this way one can ascertain how closely associated some subset of clusters may be with the distinguished clustering. If desired, clusters on non-distinguished sets can be repeated with subsets of the variables that individually evince high association with the distinguished-set clustering.

This process is presented in a diagram in Figure 2.

Green and Carmone (1977) used the log-linear models and the likelihood ratio chi-square tests of independence to conduct segment congruence analysis. This work was a continuation of an earlier paper by Green, et al (1977) in which they advocated the use of log-linear models in market structure analysis⁴.

In Green and Carmone's approach, the log-linear model is first utilized to determine the presence (or absence) and

⁴ Their work draws upon work by Bishop (1969) and Bishop, et al (1975), Goodman (1971), and Haberman (1972, 1979), who developed and used log-linear models in contingency table analysis.

the degree of mutual association among the segmentation variables in a multidimensional contingency table. Next, in the case that a distinguished base is selected a priori, the logit model is utilized to address predictability of this variable by the remaining segmentation variables.

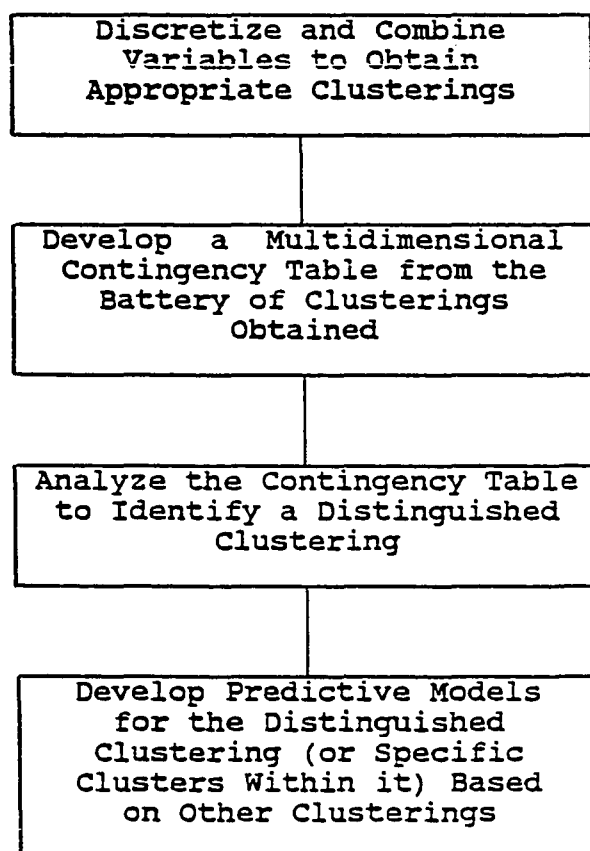


Figure 2. Segment congruence analysis process as proposed by Green and Carmone (1977).

Segment congruence analysis is, in fact, discrete multivariate analysis in the context of selection and analysis of market segmentation variables. As such, this concept need not be limited to the particular analyses introduced by Green and Carmone (i.e., the overlap, the contributions of individual variable to the overlap, and predictability of the distinguished segmentation variable by other segmentation variables.) Some instances of additional questions follows.

1. Identification of the distinguished segmentation variable can be based on several different techniques, not all of which would necessarily identify the same variable as being central.
2. It is more theoretically sound to include exogenous descriptor variables to address predictability of the distinguished base and related questions. In traditional market segmentation, the base (or segmentation variable) is identified first and, then, a group of independent variables are sought to produce profiles of the individuals within segments.
3. Assessing the order of contributions of a large number of descriptor variables would enable the researcher to select only variables which are significant in their explanatory power.
4. Finally, structure analysis has been identified as a useful area of inquiry for marketing research (particularly in descriptive analysis of market segments (Clogg and Munch 1984). A method of structural modeling of the segmentation and descriptor variables would, therefore, be well justified.

These are some additional research questions which will be addressed in this study, along with the research

questions raised by Green and Carmone (1977). Van Auken and Lonial (1984, pp.15-16) state that:

...selection of "the" superior segmentation base may have to be based on strategy and pragmatics..., rather than on a simplistic quantitative interpretation based on log-linear modeling. Still, the application of such modeling may be a prerequisite for [further] studies... Given associations, other issues may prevail, such as the prospect of combining segmentation bases, as well as the usage of a logit analysis on a distinguished segmentation base.

Admittedly, in many cases, the choice of segmentation variables is, indeed, dependent on the "expert's" understanding of the industry and its environment. However, it is through explorative studies such as segment congruence analysis that experts gain this understanding. In the explorative process of segment congruence analysis, it is reasonable to attempt to answer as many relevant questions as possible, because the incremental cost is small.

The Role of Contingency Table Analysis

Contingency table analysis is the building block of segment congruence analysis in the context of the two main methods discussed in this study⁵. Green (1977) mentioned

⁵ The history of contingency table analysis includes work by Pearson (1904), Fisher (1925), Yule and Kendall (1937), Wilks (1943), Cramer (1946), Lazarsfeld (1942), Rao (1952), Mitra (1955), Roy and Kastenbaum (1955), Kullback (1959), Friedlander (1961), Darroch (1962), Goodman (1963 thru 1978), Bishop (1969), and many others from various fields. This effort has resulted in a rich literature in contingency table analysis.

the use of multidimensional contingency table analysis as a general approach to multivariate segmentation studies.

Even though cross-tabulation of data in the form of a contingency table is a common activity in marketing research, the applications of multivariate analysis to contingency tables have been relatively scarce (Green, et al 1977). Most practical applications using contingency tables are limited to simple cross tabulations and measures of association⁶.

Contingency table analysis for market segmentation can be performed sequentially (i.e., to address multivariable problems) or simultaneously (i.e., to address multivariate problems). The purpose of analysis can be descriptive, where interrelationships are investigated symmetrically (i.e., without a notion of causality) or predictive, where the analysis are asymmetric (i.e., causal modeling). Following are brief discussions of sequential and simultaneous contingency table analyses.

Sequential Contingency Table Analysis. These techniques generally use some method of partitioning the contingency table sequentially such that some measure of dispersion within each partition is minimized. Generally,

⁶ A sample of some of the more sophisticated applications in market segmentation and related analyses includes the work by Myers and Nicosia (1967), Green, et al (1976 and 1977), Green and Carmone (1977), Rao and Solgaard (1977), Rao and Winter (1978), Russell and Rayan (1979), Fritzsche (1981), and Clogg and Munch (1984).

such analyses begin by developing a contingency table which includes, as its dimensions, a criterion variable and a number of descriptor variables. Then, the contingency table is partitioned on the descriptor variable which is deemed to give the best partitioned subtables.

Each of these subtables are further partitioned based on the best descriptor variable(s) at that level, etc. For instance, the automatic interaction detection (AID) techniques partitions the data by selecting descriptor variables which minimize the variances within subtables (Green 1978)⁷.

Some of the related approaches to AID include THAID (Morgan and Messenger 1973), Multiple Classification Analysis (MCA) (Andrews, et al 1969), Multivariate Nominal Analysis (MNA) (Andrews and Messnger 1973), and Chi-square Automatic Interaction Detection (CHAID) (Perreault, Jr. and Barksdale, Jr. 1980). These techniques have been used in marketing research frequently⁸.

This techniques for contingency table analysis are invaluable. However, like most other data analysis techniques, the very property that makes them so useful in

⁷ An information theoretic method can be developed which is completely analogous to AID, the only difference being the fact that here the entropy of each subtable is minimized.

⁸ Some examples include Armstrong and Anders (1970), Assael (1970), Martin and Wright (1974), Peters (1970), Newman and Staelin (1971), Bernhardt and Kinnear (1976), and Green (1978).

certain types of analysis limits their usefulness in certain other applications.

It should be kept in mind that the sequential partitioning methods are multivariable procedures, but not multivariate ones. All of the variables are not considered simultaneously, but rather, they are considered sequentially. Thus, a single optimal solution is not guaranteed. The partitioning which occurs in subsequent stages may be influenced by splits that have already been made. (Perreault and Barksdale 1980).

Although these techniques can be a very helpful means of reducing complexity of relationships in a large data base, and do so in a way that facilitates communication of the results, one should remember that the higher order relations will be overlooked (Perreault and Barksdale 1980). This may severely distort the analysis and yield misleading results.

Simultaneous Contingency Table Analysis. This type of analysis is not only concerned with the direct (first-order) interactions between variables but also with relationships which are a result of higher-order interactions between multiple variables. Three variables may have significant two way relationships but insignificant three way (or second-order) correlations (LaGrace 1974). In this case the researcher is able to exclude the three way interactions and

produce a simpler model for description or prediction of the essential variables.

The analysis of interactions in contingency tables have been performed through several methods. These techniques include weighted least squares, dummy variable analysis, graph theory, general probability and chi-square analysis, log-linear models, and information theory⁹.

The most relevant techniques among the above will be discussed throughout the remainder of this review in the following manner. First, the general log-linear/logit approach will be discussed. Next, the information theoretic approach will be covered. Then, other techniques which are particularly suitable for contingency table analysis will be explored.

THE LOG-LINEAR/LOGIT APPROACH

Green and Carmone (1977) proposed the general log-linear and the logit models for segment congruence analysis. This work was a continuation of an earlier paper by Green,

⁹ For excellent discussions of these techniques, see the following references: weighted least squares (Grizzle, et al 1969, and Theil 1970), dummy variable analysis (Miller and Erickson 1974, and Andrews, et al 1969), graph theory (Davis 1975) general probability and chi-square analysis (Bartlett 1955, Simpson 1951, and Lewis 1962), log-linear models (Goodman 1963, 1964a, 1964b, 1964c, 1964d, 1965a, 1965b, 1965c, 1969a, 1969b, 1972, Bishop, et al 1969, and 1975, and Haberman 1972 and 1979), and information theory (Klir 1976, Cavallo and Klir 1979, 1981, Conant 1974, Broekstra 1978, 1981, 1982, Krippendorff 1979, 1981, and Abrahamse and van Bueren 1980).

et al (1977) in which they advocated the use of log-linear models in market structure analysis. Their work draws upon work by Bishop (1969) and Bishop, et al (1975), Goodman (1971a), and Haberman (1972, 1979), who introduced log-linear models as a tool for contingency table analysis. A brief description of log-linear and logit models appears in Appendix I.

Relevant Applications of the General Log-Linear Model

In Green and Carmone's approach, the log-linear model is first utilized to determine the presence (or absence) and degree of mutual association among the segmentation variables in a multi-dimensional contingency table. Data collected in a 1975 survey of 534 respondents' views about certain aspects of their automobile insurance was used by Green and Carmone. This data was subsequently analyzed through factor analysis (in order to combine certain related variables into single variables) and (on an a priori basis) four segmentation variables were identified. These segmentation variables included "insurance supplier," "psychographic," "image," and "demographic" variables.

Then the following hypotheses were tested through the general log-linear model:

- a) All four clusterings are mutually independent.
- b) Insurance supplier is independent of the other three segmentations.

- c) Psychographics are independent of the other three segmentations.
- d) Image is independent of the other three segmentations.
- e) Demographics are independent of the other three segmentations.

The general log-linear independence model (see Appendix I) was then utilized to demonstrate the following results:

1. The four segmentation variables exhibit significant mutual association (at the alpha risk level $\leq .05$).
2. None of the four segmentations is independent of the other three.

Then, they used the drop in the chi-square between the mutual independence model and that of conditional independence models to conclude that the "image" base demonstrates the highest association with the remaining three bases. Therefore, they suggested that if one were selecting bases on a posteriori basis, the image base would be the most attractive candidate.

Next, they utilized the logit model to address the predictability of this variable by the remaining segmentation variables. This will be further discussed in the next section.

Van Auken and Lonial (1984) used Green and Carmone's approach to analyze different segmentation bases for dog

food. Their study involved a convenience sample of ninety-four dog food purchasers (This would, of course, raise questions regarding representativeness of the sample and reliability of the results, but, as an application of segment congruence analysis methodology, it is noteworthy.) Sixteen attitudinal variables, eleven benefit variables, and five problem concerns were measured, and usage data was also collected on each respondent

Each group of variables for attitudes, benefits, and problems were combined into dichotomous clusterings (This, further, raises issues concerning spurious or suppressive results created through artificial collapsings of classes of variables). For this purpose, they used factor analysis and a minimum-variance hierarchical clustering algorithm.

Using the same four hypotheses as Green and Carmone's, they concluded that the four segmentation bases attitudes, benefits, problems, and usage, are mutually associated (In reality, in spite of their claim that all four variables were mutually associated, the hypothesis of mutual independence is rejected even if only two of the variables are significantly associated.) Then using the one-variable independence log-linear models, they concluded that the problems base was independent of the other three, but attitudes, benefits, and usage were not independent of the others.

Then, by calculating the differences between the chi-square of the mutual independence model and the one-variable independence models, they concluded that the benefits base was the distinguished base. But they also used the USAGE variable for further analysis, regarding predictability, using the main-effects logit model. They further interpreted the logit model to prioritize the effect of the remaining variables. Benefits, attitudes, and problems contributed to explaining usage, in the order given here.

Other applications of the general log-linear model include Goodman (1970). The first application describes the general approach to estimation and hypothesis testing for two, three, four, and five variables (Goodman 1970). It also reanalyzes data pertaining to five dichotomized variables that classify individuals according to their information sources (e.g., newspapers, radio, books, etc.) and knowledge of cancer. He demonstrates that the log-linear models fit the data better than those obtained in earlier published analysis of the same data.

Another application by Goodman (1971b) illustrates stepwise procedures and estimation methods for building models of multiple classifications. These procedures are somewhat analogous to the classical stepwise regression methods for adding terms (forward selection) and deleting terms (backwards elimination) for qualitative models. This application gave rise to the now well-known hierarchical

log-linear models. These methods are illustrated by Goodman with a classification of 1008 consumers according to their preference for laundry detergents, previous use of a certain brand, water temperature, and water softness.

In another application, Goodman (1972) shows how recursive and nonrecursive path diagrams and more general systems can be formulated and tested using the general log-linear model. This approach is illustrated in a reanalysis of data from Wilner, et al (1955) study of the contact hypothesis, which states that the close proximity of whites and blacks in integrated housing increases favorable interracial sentiments through indirect effects by other variables such as contact and local norms.

Relevant Applications of the Logit Model

Predictive models for nominally scaled dependent variables have traditionally been based on the techniques with assumptions which are more appropriate for interval data. The traditional assumptions underlying regression analysis are violated when the dependent variable is categorical (Goodman 1978). If only the independent variables are nominally scaled, techniques for quantitative data, such as regression analysis, can be utilized by creating dummy variables which partition the nominal variable(s) into a set of sequential binary variables. Aside from the fact that in most cases it becomes extremely

cumbersome to interpret the regression results, the multicollinearity among these variables cannot conveniently be attributed to specific sets of variables.

Goodman (1972), building upon earlier work by Bishop (1969) and others, proposed a modified multiple regression approach for analyzing data of this kind. This approach uses the "logit model," a special case of the general log-linear models, and the maximum likelihood estimation procedure instead of least squares. For a description of the logit model, refer to Appendix I.

To illustrate the logit model, Goodman (1972) reanalyzed the famous study of the American Soldier by Stouffer, et al (1949), Coleman (1964), Zeisel (1968), and Theil (1970). Stouffer, et al (1949) and Coleman (1964) used regression analysis and analysis of variance to analyze this data. Coleman admits that the model differs from the actual in certain systematic ways. The variances of the estimates of the main effects would have been smaller had they used a more efficient method estimation method such as the maximum likelihood estimation methods (Goodman 1972). Zeisel (1968) and Theil (1970) used the weighted logistic-curve-transformation regression model to analyze the same data. Again, the estimates by these models have somewhat larger variances than the maximum likelihood estimates by Goodman (Rao 1965).

Goodman (1978) presented comparisons between Goodman's logit model and the dummy-variable regression model. He concluded that, though both models yield meaningful estimations, the logit model fits the data (the American Soldier data) substantially better than the corresponding additive probability models.

Chi-square measure is used to evaluate logit models. Obviously, small values of chi-square would be sought in order to establish that the logit model accounts for the variations in the data. Haberman (1982) proposed two additional measures (i.e., a measure of dispersion and one of concentration) as criteria for evaluating logit models. He gave the entropy measure as a measure of dispersion between the actual and estimated frequencies and the squared Euclidean distance between them as a measure of concentration. These measures are presented below.

$$(1) \quad H(f, F) = \sum_i f_i \log(f_i / F_i)$$

$$(2) \quad C(f, F) = \sum_i (f_i - F_i)^2 \quad ,$$

where

f_i = Actual frequency in i ($i=1, 2, \dots, N$)

F_i = Estimated frequency in i ($i=1, 2, \dots, N$)

$H(f, F)$ = Measure of Dispersion

$C(f, F)$ = Measure of Concentration

Applications of the logit models in marketing have been extremely rare. Referring back to the work by Green and his colleagues; Green, et al (1977) advocated the use of these techniques in marketing. Applying the logit model to the insurance company data, Green and Carmone (1977) assessed the predictability of an a priori selected distinguished base (i.e., the insurance supplier variable). They noted that "the objective of logit analysis is to construct a model for predicting the probability of membership in the segment representing the sponsoring firm's customers, given knowledge of the respondent's membership in the segments developed from the other three bases: psychographics, image and demographics." (Green and Carmone 1977, p.220).

Other applications of the general log-linear and logit models in marketing include consumer resource allocation, (Batsell 1980), individual and aggregate consumer brand choice (Batsell and Lodish 1981, Chapman and Staelin 1982, Currim 1981 and 1982, Gensch and Recker 1979, Jones and Zufryden 1980, Punj and Staelin 1978), marketing data analysis in general (Flath and Leonard 1979, Silk and Urban 1978), consumer response models (Hauser and Urban 1977), information load and information processing (Malhotra 1982a and 1982b, Malhotra, et al 1982).

THE INFORMATION THEORETIC APPROACH

A concept which is fundamental in general system theory and methodology and in system theoretic data analysis in particular is the concept of a system. "In general, a system is an abstraction distinguished on an object by an observer, which reflects the interaction between the observer and the object" (Klir 1986, p. 267). In the data analytic sense, this abstraction is conceived as a set of variables together with a characterization of relationship, dependency, correlation, or any other expression of a constraint among the variables.

Constraint analysis as proposed by Ashby (1964), and later developed by Klir and his colleagues, evaluates a system in terms of the information contained in its given subsystems. "Information is considered here in terms of the reduced uncertainty it produces in the overall system inferred (Klir 1986, p. 269)."

To make this principle [of the information content within a system] operational, a unique and well justified measure of uncertainty must be determined for the mathematical formalism employed (Klir 1986, p. 269).

Such measure exists in statistical information theory. Herniter (1973) gives the following statement about information theory's measure of uncertainty:

The great advance provided by information theory lies in the discovery that there is a unique, unambiguous criterion for the "amount of uncertainty" represented by a discrete probability distribution, which agrees with our intuitive notions that a broad distribution represents more uncertainty than does a sharply peaked one, and satisfies all other conditions which make it reasonable.

Statistical information theory draws its foundations from "entropy" in thermodynamics. Entropy, as used by Shannon (Shannon and Weaver 1949), refers to uncertainty associated with a signal transmitted through a channel of communication. The amount of "information" in a message is defined as the reduction of the uncertainty achieved by receiver of the message. "It has been proven in several alternative ways that Shannon entropy is the only function that possesses all known properties for a probabilistic measure of uncertainty." (Klir 1986, p. 269)

An extensive body of literature has been devoted to the study of the relationship between entropy and information (Stumpers 1953; Bartlett 1955; Brillouin 1956; and Cherry 1952; have produced extensive bibliographies concerning this relationship.)

The work by McGill (1954), Kullback (1959), Garner (1962), Theil (1967), Ku and Kullback (1968), and many others, has further developed information theory and its application to social sciences.

Applications of the information theoretic models to marketing include the following: Bernhardt and MacKenzie (1968), Herniter (1973 and 1974), Bass (1976). A brief summary of the concept and some extensions of information theory appears in Appendix II.

This study employs Shannon's (Shannon and Weaver 1949) entropy measure as a measure of uncertainty (and hence the information that can be acquired when this uncertainty is reduced in a system) to assess discrete multivariate systems. It will be demonstrated that the data analysis properties of information theory, also referred to as uncertainty analysis (Conant 1974) and constraint analysis (Broekstra 1978, 1981), provide a powerful vehicle for assessing interrelationships and the structure of a system.

Some Important Properties of Information Theory

Recent developments in information theoretic methods for constraint analysis, reconstructability analysis, structural modeling (Conant 1974, Klir 1976, Cavallo and Klir 1979, 1981, Broekstra 1981, and Krippendorff 1981) have promised an approach to contingency table analysis different from and potentially superior to the general log-linear and logit methods. These methods enable the researcher to assess the structure of the data with respect to their mutual association, the strength of each variable's contribution to the total mutual association, and

predictability of each variable based on the other variables.

This is done by utilizing the decomposability property of the information theoretic measures which enables assessment of total interactions as well as partial interactions among variables. The present undertaking will demonstrate that segment congruence analysis can be performed using the information theoretic approach.

The most important properties of the entropy measure (Horowitz and Horowitz 1976) include the following:

- 1) As the probabilities of possible states tend to equalization, the entropy of the system approaches its maximum.
- 2) Between two systems at maximum entropy the one with a larger number of states has a higher entropy measure.
- 3) Entropy of a system is between zero and the logarithm of the number of possible states of the system.
- 4) It is possible to calculate all conditional entropies including the measures parallel to the Bayesian posterior probabilities.
- 5) The entropy measure is fully decomposable, enabling the calculation of entropies within and between groups in a manner similar to ANOVA.

It has been shown that information theoretic data analysis and the analysis of variance are analogous. Uncertainties can be estimated directly from variances in certain cases, using the relationships between the two. (Garner and McGill 1956, Attneave 1959). The decision as to

which of the techniques to use depends partly on the properties of the criterion variable.

Only uncertainty analysis may be used with non-metric criterion variables since uncertainties are dimensionless (using no metric), however, uncertainty analysis has a generality which may make it useful even when variances can be computed. (Garner and McGill 1956, p. 219)

Considering the fact that research in social sciences (including marketing) uses qualitative data extensively, information theoretic analysis can be applied in this area. This was the primary motivation for work in information theory by a great number of researchers.

Information Theoretic Definitions of a Leading Part

The concept of leading part was initially introduced by von Bertalanffy (1968) in connection with tendency of systems for segregation and/or centralization. He defined this concept in the following manner:

Suppose that the coefficients of one element, P_s , are large in all equations [of a system of differential equations, as a model of a system under investigation,] while the coefficients of the other elements are considerably smaller or even equal to zero. ... We may call the element P_s a leading part, or say that the system is centered around P_s . (von Bertalanffy 1968, p.71)

The significance of a leading part is in that a small change in this part (e.g., element, variable, or group of variables) may cause a significant change in the whole

system. In information theoretic data analysis, similar to the other descriptive data analysis methods, much attention has been given to identifying one (or a set of) most significant variable(s).

Several schemes for information theoretic definition of the leading part have been developed by Zwick (1979), which will be further extended and presented briefly in this section. (The reader unfamiliar with information theory should review Appendix II at this point). The information theoretic leading part can be defined according to, at least, four perspectives:

1. The part which makes the highest contribution to the transmission.
2. The part which makes the highest contribution to the joint variability (systematic entropy) in the system.
3. The part which makes the highest contribution to the overall uncertainty in the system.
4. the part which has the greatest number of interactions with the other variables.

The mathematical expressions for these four schemes are developed in Chapter 3 (Research Questions, Methodology, and Procedures), in the context of segment congruence analysis. Those expressions can easily be generalized to include all multivariate systems. Therefore, they will not be repeated here.

Statistical Significance of Informational Measures

Based on Wilks' (1935) work and Miller's (1953) proof, informational measures are known to be asymptotically chi-square distributed. This work facilitated the application of information theory to social science problems, where sampling is prevalent, and the analyses are performed to assess the behavior of the population from which the sample is drawn.

Significance Test for the Transmission Measure. With N denoting the sample size, the exact likelihood ratio chi-square value for the information theoretic measure of transmission is $2.N.T.LOG(2)$. Attneave (1959), using the then much more common χ^2 (Pearson's chi-square) states that this relationship is an approximation. However, in Table II (Part a) derivation of likelihood ratio chi-square from transmission is presented and it is shown that this relationship is exact. Therefore, using information theoretic transmission measures provides the researcher with tests of independence, using the likelihood ratio chi-square with degrees of freedom given by Broekstra (1981, p.51) and presented below:

$$(3) \quad df(T) = - \sum_i (|s(v_i)| - 1) + |A| - 1$$

In equation 3, i = number of variables in T , $s(v_i)$ = number of states (classes) in variable i , and $|A|$ = number of aggregate states (cells) in the contingency tables.

Significance Test for the Entropy Measure. In principle, the chi-square tests are capable of entertaining any null hypothesis with respect to the distribution (or generating function) of the expected frequencies. Blalock (1972, p.275) states that: "The chi-square test is a very general test that can be used whenever we wish to evaluate whether or not frequencies which have been empirically obtained differ significantly from those which would be expected under certain set of theoretical assumptions."

The chi-square test of independence assumes that the expected (cell) frequencies are products of the marginal frequencies. This (independence) is an implicit assumption in the calculation of the transmission measure.

In order to test the entropy measure, one can calculate the chi-square value using the empirically observed frequencies against a uniform distribution. The null hypothesis would then assess the departure from the uniform distribution (i.e., maximum entropy) for the observed frequencies. The calculation of the likelihood-ratio chi-square for measures of entropy is presented in Table II (Part b)

As all information theoretic measures can be converted to entropy terms (see, for instance, Krippendorff 1979),

this procedure provides a test of significance for all such measures. The literature search, conducted as part of this research, has not found any previous work utilizing this relatively trivial but useful conclusion.

Degrees of freedom in this case are calculated as $n-1$, where n is the number of cells in the contingency table (the only constraint being the sum of relative frequencies (i.e., probabilities must) be one. For the more complex measures (e.g., systematic entropy), by analogy to the degrees of freedom for complex transmissions (Broekstra 1981), the algebraic sum of the degrees of freedom would produce the total number of degrees of freedom.

The derivation in Table II (Part b) has an implication that can prove to be quite useful. By performing the appropriate linear transformation on H , one can obtain the likelihood-ratio chi-square for the entropy measure. Though, the likelihood-ratio chi-square measure has been used extensively for the transmission based measures, it has not been used for statistical assessment of information theoretic measures other than transmission, such as entropy. This derivation shows that L^2 for the transmission measure can be used to test departure from independence while L^2 for the entropy measure can be used to test departure from maximum entropy.

TABLE II

RELATIONSHIP BETWEEN LIKELIHOOD-RATIO CHI-SQUARE
AND THE INFORMATION THEORETIC MEASURES

(Part a) Calculation of the likelihood-Ratio chi-square for the transmission measure

$$L^2 = 2 \sum O \log(O/E),$$

where, $O = f_{ij} = n_{ij}P_{ij}$, and $E = F_{ij} = n_{ij}P_{i.}P_{.j}$

$$= 2 \sum_i \sum_j n_{ij} P_{ij} \log(n_{ij} P_{ij} / n_{ij} P_{i.} P_{.j})$$

$$= 2N \sum_i \sum_j P_{ij} \log(P_{ij} / P_{i.} P_{.j})$$

$$= 2N \sum_i \sum_j P_{ij} (\log P_{ij} - \log P_{i.} - \log P_{.j})$$

$$= 2N (\sum_i \sum_j P_{ij} \log P_{ij} - \sum_i P_{i.} \log P_{i.} - \sum_j P_{.j} \log P_{.j})$$

$$T = \sum_i \sum_j P_{ij} \log_2 P_{ij} - \sum_i P_{i.} \log_2 P_{i.} - \sum_j P_{.j} \log_2 P_{.j}$$

therefore;

$$L^2 = 2N \log(2) (T)$$

(Part b) Calculation of the likelihood-Ratio chi-square for the entropy measure

$$L^2 = 2N \sum_i \sum_j P_{ij} \log P_{ij} / (1/A)$$

where

A = Total number of states (i.e., number of cells in the contingency table)

$$= 2N \sum_i \sum_j P_{ij} \log P_{ij} + 2N \sum_i \sum_j P_{ij} \log (1/A)$$

$$= 2N \log(2) \sum_i \sum_j P_{ij} \log_2 P_{ij} - 2N \log(A) \sum_i \sum_j P_{ij}$$

$$= - 2N \log(2) (H) + 2N \log(A)$$

$$= - 2N \log(2) [H - H_{MAX}]$$

Information Theory as a Tool for Segment Congruence Analysis

Information theory strengthens segment congruence analysis in several critical ways. (1) It readily provides a measure of the strength of association as well as addressing the question of statistical existence of the association. (2) The leading part identification schemes provide a tool of identifying several candidates for segmentation variables. (3) It avails the structure analysis techniques, which are very well developed using the information theoretic ideas, to researchers in marketing and other social sciences. Following is a brief discussion of these features and their critical role in data analysis.

Information Theoretic Measures of Association as a measure of overlap in multivariate data. As will be seen in the methodology chapter, the general log-linear approach hypothesize a model for the data (e.g., all variables are independent). This model is then used to generate the expected cell frequencies, using the functions of the appropriate marginals, or in some cases (when no direct estimations are possible-- to be explained later) an iterative process. Next, the expected cell frequencies, generated based on the model, are compared with the actual cell frequencies to draw statistical inference regarding the goodness-of-fit of the model.

But this only assesses the existence (or lack of existence) of a particular hypothesized relationship in the data. Blalock (1972, p.291) states the following:

We have set up null hypotheses to the effect that [e.g.,] there is no relationship and have then tried to reject these hypotheses. But just how much have we accomplished when we are able to reject? We refer to a relationship as being statistically significant when we have established, subject to the risk of a type I error, that there is a relationship between the two [or more] variables. But does this mean that the relationship is significant in the sense of being a strong relationship or an important one? Not necessarily. The question of the strength of a relationship is a completely different question from that of whether or not the relationship exists.

Although many measures of association are available for assessing the strength of relations, the information theoretic measures of association (e.g., transmission and systematic entropy) have the following advantages:

- a) They have definite minimum and maximum values.
- b) There exists a direct formula which produces the likelihood-ratio chi-square for each measure.
- b) Different models can be hypothesized using information theoretic models.

No other technique (e.g., the general log-linear models or the chi-square alone) has all of the above properties. The log-linear approach only provides a scheme of generating the cell frequencies, and likelihood-ratio chi-square alone is not nearly as flexible (nor as established) in model generation.

Use of the Concept of Leading Part in the Identification of Alternative Candidates for the Distinguished Segmentation Variable. There exists a great deal of controversy concerning whether the post-hoc (or even the hybrid) segmentation schemes identify meaningful market segments. (e.g., Green 1977, Wind 1978) This is partly due to the fact that most clustering-based segmentation methods identify a single variable (or clustering) as the segmentation base, then, proceed to assess their predictability through the background variables.

The four definitions of leading part enable the researcher to, first, identify several candidate segmentation variables (or clusterings). Then, through the assessment of predictability of these candidate variable, a distinguished variable (or clustering) is identified. This is obviously a critical advantage of using the proposed information theoretic approach.

Advances in Information Theoretic Structure Analysis Useful in Segment Congruence Analysis. As a descriptive family of techniques structural modeling is invaluable in any multivariate analysis context. For instance, one of the reasons for clustering-based techniques' inability to produce meaningful market segments can be attributed to the lack of a thorough understanding of the structure and strength of interrelationships among the variables in the models analyzed.

Researchers in general system methodology have consistently been at forefront of structural modeling techniques. Constraint analysis, dependency analysis, reconstructability analysis, etc. are familiar terms in general system methodology. The primary tool used for these analysis techniques is information theory.

The log-linear based hierarchical modeling or the general probability models are still in infancy. The methods used to generate and assess models parallel those of information theoretic approach, however, information theory is more established and defined.

Information theory, therefore, offers a valuable family of techniques to the researcher in marketing, which improves his/her understanding of the complex multivariate systems. This underutilized approach can enable the marketing practitioners to make better-informed decisions.

THE GENERAL PROBABILITY AND CHI-SQUARE APPROACH

In addition to log-linear and information theoretic methods, a variety of other techniques have been used to address mutually associated variables. Among these methods, only the general probability and chi-square analysis warrants inclusion in this review. Two reasons justify this inclusion: 1) this approach has occasionally been used to analyze contingency tables directly (Goodman 1971a, 1974,

Dempster, et al (1977), Clogg and Munch 1984), and 2) the chi-square measure is used in both log-linear and information theoretic approaches to assess statistical significance for different models.

The general probability approach primarily deals with calculations of joint and conditional probabilities of occurrence of mutually dependent events. Principles of this approach are present in (and used by) both the general log-linear models and the information theoretic models. Some recent work in this area is attributed to Goodman (1974), Dempster, et al (1977), and Clogg and Munch (1984), among others.

Clogg and Goodman (1985) developed the simultaneous latent structure analysis model in general terms. This model is a probability based model which searches for unobservable (latent) variable(s) underlying the data. These models could also be used for studying structural models by searching for the observable variable(s) that are best determined by the other variables, however, no applications of this was found.

Chi-Square Test of Independence. In general terms, independence means that the probability of one event is not affected by occurrence or non-occurrence of another event. Variables in a contingency table are independent if the probability of occurrence of each composite class (e.g., P_{ijk}) is equal to the product of the relevant marginal probabilities (i.e., $P_{i..} * P_{.j.} * P_{..K}$, where, $P_{i..} = \sum_j \sum_k P_{ijk}$, etc.).

To test the hypothesis that an observed cross-classification fits the independence model, the estimated expected cell-frequencies are calculated from the marginal frequencies:

$$(4) \quad F_{ijk} = n * (f_{i..}/n) * (f_{.j.}/n) * (f_{..K}/n)$$

These expected values are then compared with the observed values using either the Pearson chi-square statistic (2), or the likelihood-ratio chi-square (3):

$$(5) \quad X^2 = \sum_i \sum_j \sum_k (f_{ijk} - F_{ijk})^2 / F_{ijk}$$

$$(6) \quad L^2 = 2 \sum_i \sum_j \sum_k f_{ijk} * \log(f_{ijk}/F_{ijk})$$

"Although less widely known than GFX^2 [Pearson X^2], the likelihood ratio chi-square has the advantage of being divisible into additive portions, much as the total sum of squares in the analysis of variance is partitioned."

(Reynolds 1977, p.8-9) "Both statistics are asymptotically distributed as chi-square, with $n - p$ degrees of freedom, where n is the number of cells in the table, and p is the number of independent parameters fitted." (DeSarbo and Hildebrand 1980, p.44)

SUMMARY

This chapter produced a review of the literature in segment congruence analysis and other relevant concepts and techniques. First, it was established that in order to perform segment congruence analysis in its intended form, a multidimensional contingency tables of segmentation variables must be developed. This highlighted the significant role of multivariate discrete analysis using contingency tables. It was then demonstrated that techniques for sequential analyses or those which assume interval dependent variables are not suitable for segment congruence analysis.

Next, simultaneous discrete multivariate analysis was introduced, followed by concepts and applications of log-linear/logit, information theoretic, and general probability methods of performing such analysis. It was seen that, in spite of great interest in the information theoretic approach, actual applications of these techniques in marketing are scarce. Considering that segment congruence

analysis seems to be a logical area of application for information theory, it is justified to use and compare this technique with other traditional methods.

Table III presents an outline of different sections of this literature review and a sample of relevant publications in each area.

TABLE III
SELECTED PUBLICATIONS IN THE AREAS CENTRAL
TO THE PRESENT STUDY

TOPIC	SOURCE	DATE
o Segment	Green & Carmone	1977
Congruence	Green	1977
Analysis	Arabie, et al	1981
	Van Auken & Lonial	1984
	Harmon, Zwick, & Hosseini	1986
o Contingency Table Analysis:		
-Log-Linear	Haberman	1972, 1979
Approach	Goodman	1963-1972
	Bishop	1969
	Bishop, et al	1975
	Green, et al	1977
	Green & Carmone	1977
	Malhotra	1982
	DeSarbo & Hildebrand	1980
	Currim	1981, 1982
	Batsell	1980
	Batsell & Lodish	1981
	Chapman & Staelin	1982
	Gensch & Recker	1979
	Jones & Zufryden	1980
	Punj & Staelin	1978
	Flath & Leonard	1979
	Silk & Urban	1978
	Hauser & Urban	1977
-Information	Klir	1976-1981
Theoretic	Cavallo & Klir	1979, 1981
Approach	Conant	1974-1982
	Broekstra	1978-1981
	Krippendorff	1976-1981
	Abrahamse & van Bueren	1980
	Bernhardt & MacKenzie	1968
	Herniter	1973, 1974
	Bass	1976
-General	Bartlett	1935
Probability	Simpson	1951
Chi-Square	Goodman	1963-1978
Analysis	Clogg & Goodman	1985
	Dempster, et al	1977
	Clogg & Munch	1984

CHAPTER III

RESEARCH QUESTIONS, METHODOLOGY, AND PROCEDURES

This chapter is composed of three parts: research questions, methodology, and description of the data used. First, a description of the research questions addressed in this study is presented. Next, the information theoretic and the log-linear/logit approaches are described in the context of segment congruence analysis is given. Finally, a description of the study and the procedures employed for selection and aggregation of variables is provided.

RESEARCH QUESTIONS

This section describes the research questions proposed for conducting segment congruence analysis and their significance in general and as applied to the data used (i.e., the residential energy survey data). When multiple segmentation variables are used, the researcher must investigate the implications of interactive relationships among these variables (i.e., mutual causal effects, strength of relations, predictability of each variable or set of

variables, reduction of the predictor-variable set, and the structure of the segments). This investigation is conducted in this study through two distinct phases, (1) analysis of the overlap, and (2) identification and analysis of the distinguished segmentation base.

Phase I: Analysis of the Overlap

This phase concentrates on the issues of mutual association among various segmentation variables. At this stage, therefore, only the segmentation variables are used. These variables (or clusterings) were identified in the PNWRES study as GENATT, general attitudes toward economy and environment, ENRATT, energy related attitudes, BEHAVE, energy consumption behavior, PERCEPT, energy (comfort-discomfort) perceptions, and KWHUSE, energy usage in kilowatt-hours. (Procedures used and the rationale for selection of these variables (or clusterings) are discussed in detail in the third section of this chapter). Five research questions were addressed:

1) Are the segmentation variables mutually associated? If yes, how can this mutual association be measured?

For instance, assuming that each consumer is assigned to each one of four bases independently, a multivariate

profile for each individual results: $X_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4})$, where X_{ik} indicates membership in segment i of segmentation variable k ($k=1, 2, 3, 4$) (Green and Carmone, 1977). Then, the question of whether these bases are mutually interdependent is useful in determining whether one will know anything about the individual's membership in other segments when his/her membership in a set of (one or more) segments is known.

The hypothesis tested for the mutual association assessment is as follows:

H_0 : All of the variables in the system of segmentation variables are independent.

H_A : Two or more of the segmentation variables are interdependent

Question #1, in the context of PNWRES study is stated as: Are the variables GENATT, ENRATT, BEHAVE, PERCEPT, AWARE, and KWHUSE mutually associated? I.e., do attitudes, behavior, perception, awareness, and energy use overlap, such that knowing about one of these, e.g., about a person's attitudes, would reduce one's uncertainty about another, e.g., energy use? If so, how would this association be measured using the general log-linear model versus the information theoretic method.

2) Which basis (or set of bases) makes the highest contribution toward the mutual association?

This question addresses a method to obtain a candidate distinguished base. The answer to this question is particularly useful when the remaining segmentation variables are "accessible." One can then designate the selected variable as the distinguished segmentation variable and the remaining variables as descriptor variables.

A sample of the hypotheses tested for contributions of individual variables to the mutual association are similar to the following:

H_0 : The variable (e.g., GENATT) does not contribute toward the total mutual association

H_A : The variable (e.g., GENATT) contributes toward the total mutual association

As applied to the PNWRES study, this research question becomes: Which one (or set of) variables GENATT, ENRATT, BEHAVE, PERCEPT, AWARE, and KWHUSE contributes the most toward the mutual association measured through question #1? Such a set would then be designated as a candidate for the distinguished base.

3) Which basis (or set of bases) makes the highest contribution toward the joint variability in the system of segmentation bases?

This question addresses a second method for identifying a candidate segmentation variable (or set of variables). In information theoretic terms, this research question differs from the previous one in that the mutual association addresses the transmission in a multivariate system, while joint variability addresses the systematic entropy.

A multivariate system may include overlapping variables for which usual tests of independence indicate statistically significant mutual association. But, since the usual association measures (e.g., chi-square, Yule's Q, phi-square, transmission measure, etc.) include higher levels of interaction more than once (i.e., two-way interactions are represented once, three-way interactions are represented twice, etc.), a statistically significant association measure does not necessarily indicate that a significant portion of the total system has been accounted for.

Krippendorff's "systematic entropy," which in principle measures the joint area between variables only once, will be utilized to assess how much of the uncertainty in the system is accounted for. This is the same as

measuring the area of uncertainty shared between variables in a multivariate system.

A sample of the hypotheses tested for contributions of individual variables to the joint variability in the system of segmentation variables:

H_0 : The variable (e.g., GENATT) does not contribute toward the joint variability in the system of segmentation variables.

H_A : The variable (e.g., GENATT) contributes toward the joint variability in the system of segmentation variables.

In the context of the PNWRES study this question investigates the contribution of each variable to the joint variability between GENATT, ENRATT, PERCEPT, AWARE, BEHAVE, and KWHUSE.

4) Knowledge of which base (or set of bases) makes the highest contribution toward the reduction of overall uncertainty in the overall system of segmentation bases?

This is a third way of identifying a candidate distinguished segmentation base. Definition of the distinguished set is analogous to von Bertalanffy's (1968) definition of the "leading part". The identification of the leading part in a multivariate system can be based on the search for the variable-set which contributes most significantly to the "systemness" of the multivariate system

(i.e., to the mutual association, or to the joint variability in the system addressed in research questions #2 and #3, above), or it can be based on the set that reduces the total uncertainty in the multivariate space (research question #4).

A sample of the hypotheses tested for contributions of individual variables to the overall uncertainty in the system of segmentation variables:

H_0 : The variable (e.g., GENATT) does not contribute toward the overall uncertainty in the system of segmentation variables.

H_A : The variable (e.g., GENATT) contributes toward the overall uncertainty in the system of segmentation variables.

Again, in the PNWRES study this question would address the identification of a third candidate for the distinguished base. The question may be stated as: Of the above variables, which one (or set) reduces the uncertainty in the system $S=(\text{GENATT}, \text{ENRATT}, \text{BEHAVE}, \text{PERCEPT}, \text{AWARE}, \text{KWHUSE})$, the most?

A fourth way of determining the leading part is to identify the variable-set which is most central in terms of pairwise interrelationships with other variables in the system. This property is explored through research question #5.

5) Which variable (or set of variables) has the greatest number of statistically significant dyadic interrelationships with other variables?

This question addresses a fourth method for identifying a candidate distinguished set. In this method the variable (or a set of variables) which interacts with a greater number of other variables is considered as being more central, and therefore, distinguished.

A sample of the hypotheses tested for the dyadic interrelationships with other variables:

H_0 : The two variables being considered are independent.

H_A : The two variables are interdependent.

Phase II: Identification and Analysis of the Distinguished Variable

This phase addresses the analysis of the candidate distinguished bases identified in phase I, and selection of a best segmentation base. In this phase, the study will proceed along two distinct paths. First, the other segmentation variables are used as descriptor variables to select and analyze the distinguished base. Second, a set of exogenous descriptor variables will be used to select and analyze the distinguished base.

This second approach, therefore, requires that a set of independent (descriptor) variables be identified.

Considering that a large number of descriptor variables may be available, it is necessary to address the issues of data reduction and reinterpretation. This can be done by assessing the data in order to identify a parsimonious set of variables which best determine the segmentation variable.

The descriptor variables identified for the PNWRES study include geographic/climatic environment, CLIMGEO, type of dwelling, TYPDWEL, the status of the resident in terms of renter/owner distinction, RENTOWN, the demographic cluster, DEMOG, size of the dwelling, DWELSIZE, and the level of insulation of the residence, INSUL. (A detailed discussion of selection and aggregation of variables which comprise each of the above variables (or clusterings) is given in the third section of this chapter.)

If it is assumed, for the sake of discussion, that, for example, KWHUSE, ENRATT, and BEHAVE have been identified as candidates for the distinguished base, the following four research questions are posed in this second phase of analysis:

1) Which one of the candidate distinguished segmentation bases, identified in phase I, is best predicted once we have the knowledge of a set of relevant independent variables?

As already mentioned, once a segment (or a set of segments) is identified, the researcher needs to know how to identify and reach the individuals within this segment. It is certainly useful to know that the high electrical consumption group is comprised of individuals who, say, reside in north side, are married, have two or more children, and live in houses that were built prior to 1960, etc. The present research question addresses this issue by searching for the candidate segmentation base best predictable given a set of demographic, socioeconomic, geographic, and/or psychographic variables. (Predictability means knowing the class of the dependent variable, given the classes of the independent variables).

Therefore, in the PNWRES study the predictability is addressed as: Which base (KWHUSE, ENRATT, or BEHAVE) is best determined (i.e., accounted for) once we have the knowledge of the descriptor variable set (CLIMGEO), (TYPDWEL), (RENTOWN), (DEMOG), (DWELSIZE), and (INSUL)?

2) How much of the uncertainty in the distinguished variable is reduced when it is dichotomized into a class (or group of classes) versus all of the other classes?

It is possible to isolate the segment which is of particular interest as one class, and all other segments as another (i.e., to form a dichotomous segmentation variable, where one of the classes is composed of the "sought" group). Then, the predictability (or reduction in uncertainty) in the new variable due to the knowledge of the descriptor variables can be investigated.

In the PNWRES study, assuming that KWHUSE (containing 5 classes) is selected as the distinguished base in question #1 of Phase II, this question may be stated as: How much of the uncertainty in the dichotomized KWHUSE is reduced with the knowledge CLIMGEQ, TYPDWEL, RENTOWN, DEMOG, DWELSIZE, and INSUL under the following dichotomization schemes?

- 1) Classes 1&2 vs. all other classes (the low-consumption group vs. others).
- 2) Class 3 vs. all other classes (the medium-consumption group vs. others).
- 3) Classes 4&5 vs. all other classes (the high-consumption group vs. others).

The information obtained through these analyses is useful in assessing predictability of individual groups.

One knows how much is known about the variable of interest given the knowledge of the descriptor variables.

3) How can one prioritize the effect of the independent variables on the distinguished set?

This question, in a way, addresses the principle of parsimony in obtaining data on the independent variables (i.e., which independent variables must be retained and which ones can be dropped). If, in prioritizing the amount of prediction made by each independent variable, it becomes evident that a variable contributes very little to predicting the dependent variable, it may be disregarded and more effort spent on obtaining data on variables with higher priority.

In the PNWRES study, under the assumption that KWHUSE is selected as the distinguished base in question #1, Phase II: Prioritize the effect of (CLIMGEO), (TYPDWEL), (RENTOWN), (DEMOG), (DWELSIZE), and (INSUL) in determining the segmentation base KWHUSE-- i.e., which one contributes the most toward the determination of KWHUSE, which one contributes second most, etcetera?

After eliminating the descriptor variables with insignificant contribution toward the segmentation base determine the answer to the following question:

4) What is the dependency structure in each class of the distinguished segmentation base?

In general system methodology, a structure is interpreted as "a maximally-simple, adequately-complete set of subrelations which taken together capture most or all of the constraints and relations exhibited by the system..." (Conant 1982, p.243). Dependency analysis as developed by Conant (1981, 1982) deals with a method of determining an acceptable structural model (called dependency structure) for a system taking into account the strength of relations among its variables. Later in this chapter, Conant's method will be described in detail.

This research question not only addresses the construction of a structural model of the data (i.e., a diagram), but to assess the strength of the direct effects in such model. Furthermore, the subset of (statistically) significant variables in the system can be interpreted as the leading part.

Finally, in the PNWRES study one would want to: Quantify the dependency structure for each of the structural models obtained in 3 above-- i.e., given each structural model find strong and weak interrelationships. This quantification is significant due to the same reasons given for the fourth research question (i.e., gain in the

understanding of the determinants of the criterion variable and their relationships).

Tables IV and V presents summary of the research questions and their significance in segment congruence analysis in Phases I and II, respectively. All of the research questions have significance and are of interest to marketing practitioners and theoreticians in general including those in the electrical utility industry (and other energy related fields).

TABLE IV

RESEARCH QUESTIONS AND THEIR SIGNIFICANCE IN
SEGMENT CONGRUENCE ANALYSIS (PHASE I)

<u>RESEARCH QUESTION #1</u>	Are the segmentation variables mutually associated? If yes, how can this mutual association be measured?
SIGNIFICANCE:	Is it possible, by obtaining information about a subset of these variables, to learn about the remaining variables?
<u>RESEARCH QUESTION #2</u>	Which basis (or set of bases) makes the highest contribution toward the mutual association?
SIGNIFICANCE:	Identifying a candidate for the distinguished segmentation set (Method 1).
<u>RESEARCH QUESTION #3</u>	Which basis (or set of bases) makes the highest contribution toward the joint variability in the system of segmentation bases?
SIGNIFICANCE:	Identifying a candidate for the distinguished segmentation set (Method 2).
<u>RESEARCH QUESTION #4</u>	Knowledge of which base (or set of bases) makes the highest contribution toward the reduction of overall uncertainty in the overall system of segmentation bases?
SIGNIFICANCE:	Identifying a candidate for the "distinguished" segmentation set (Method 3).
<u>RESEARCH QUESTION #5</u>	Which variable (or set of variables) has the greatest number of statistically significant dyadic interrelationships with other variables?
SIGNIFICANCE:	Identifying a candidate for the "distinguished" segmentation set (Method 4).

TABLE V
RESEARCH QUESTIONS AND THEIR SIGNIFICANCE IN
SEGMENT CONGRUENCE ANALYSIS (PHASE II)

<u>RESEARCH QUESTION #1</u>	Which one of the candidate distinguished segmentation bases, identified in phase I, is best predicted once we have the knowledge of a set of relevant independent variables?
SIGNIFICANCE:	To select a "distinguished" segmentation basis (or a set of bases) amongst the candidates obtained in questions 2, 3, 4, and 5 of Phase I.
<u>RESEARCH QUESTION #2</u>	How much of the uncertainty in the distinguished variable is reduced when it is dichotomized into a class (or group of classes) versus all of the other classes?
SIGNIFICANCE:	This will show which class (or group of classes) of the distinguished variable is best represented by the independent variables in the system.
<u>RESEARCH QUESTION #3</u>	How can one prioritize the effect of the independent variables on the distinguished set?
SIGNIFICANCE:	To rank order the contributions of the independent variables to explaining (reducing uncertainty in) each dependent variable.
<u>RESEARCH QUESTION #4</u>	What is the dependency structure in each class of the distinguished segmentation base?
SIGNIFICANCE:	What is the ordinality of relationships among variables?

METHODOLOGY

This section includes four parts. The first part describes the information theoretic expressions used in assessing each of the research questions. The second part entails a brief description of the computer programs used to perform the information theoretic analysis. The third part describes, where applicable, the log-linear and logit models used to assess the research questions. And finally, the fourth part discusses the reasons for using the information theoretic methodology as compared with the more commonly used log-linear and logit models.

Information Theoretic Methodology

As mentioned earlier, Appendix II introduces the basic concepts in information theory and its extension to analysis of dependence and independence in a multivariate system and structural modeling. Information theoretic methods can be used to address the research questions introduced in the previous part.

Phase I: Analysis of the Segmentation Variables. This phase addresses the assessment of mutual association among the segmentation variables, and then, discusses four methods of identifying candidate segmentation variables.

Research Question 1. Information theoretic entropy transmission can be used to assess interrelationships among a number of segmentation variables. To measure the degree of association among variables A, B, C, and D, the transmission measure, $T(A:B:C:D)$, is computed using equations (7) and (8):

$$(7) H(A,B,C,D) = - \sum_i \sum_j \sum_k \sum_l P(A_i, B_j, C_k, D_l) \log_2 P(A_i, B_j, C_k, D_l)$$

$$(8) T(A:B:C:D) = H(A) + H(B) + H(C) + H(D) - H(A,B,C,D)$$

where $i=1,2,\dots,n_A$, $j=1,2,\dots,n_B$, $k=1,2,\dots,n_C$, $l=1,2,\dots,n_D$, and n_A , n_B , n_C , and n_D are numbers of classes of variables A, B, C, and D, respectively.

This addresses the first research question in phase I (i.e., the mutual association among the segmentation bases). Transmission in a multivariate system measures the amount of overlap between the uncertainties of the variables. Additionally, it measures this overlap with respect to the ordinalities of the overlaps: Those parts that overlap once (i.e., demonstrate two-way interactions) are represented once, those with three-way interactions are represented twice, those with four-way interactions are represented three times, etc.

Figure 3 demonstrates this property spatially. In terms of the areas numbered in the Venn diagram in Figure 3, transmission is represented as:

$$T(A:B:C:D:) = H(A) + H(B) + H(C) + H(D) - H(A,B,C,D)$$

$$\begin{aligned} (2,4,5,5,6,7,7,8,8,8,9,9,10,11,11,12) = \\ (1,2,4,5,7,8,10,11) + (2,3,5,6,8,9,11,12) + \\ (7,8,9,10,11,12,13,15) + (4,5,6,7,8,9,13,14) - \\ (1,2,3,4,5,6,7,8,9,10,11,12,13,14,15). \end{aligned}$$

Note that overlapping areas 2, 4, 6, 10, 12, and 13 are included once in the total transmission; 5, 7, 9, and 11 are included twice, and 8 is included three times.

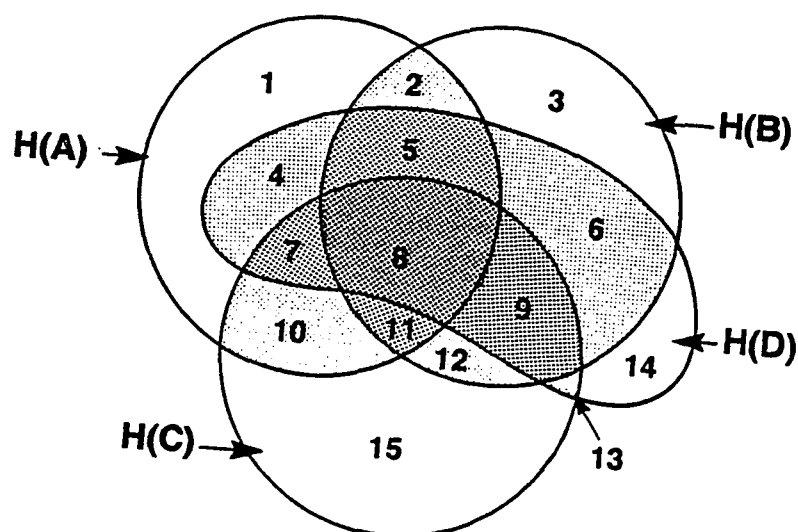


Figure 3. Spatial representation of transmission.

As previously mentioned, transmission measure approximates a chi-square distribution (Attneave, 1959) with the likelihood ratio chi-square equal to $2 \cdot \log(2) \cdot N \cdot T$. The degrees of freedom are computed based on the formula given

by Broekstra (1981) and reproduced in Chapter II (see Page 43).

Transmission-based measures of association have been used widely in general system methodology (Broekstra, van Bueren, etc.). A normalized measure of transmission can be obtained by computing the ratio as shown in Equation (9):

$$(9) \quad T_{\text{NORM}} = \frac{T}{T_{\text{max}}} = \frac{T}{\sum_i H(X_i) - \text{Max}\{H(X_i)\}}$$

This measure has a minimum of zero and a maximum of 1. These are desired properties in evaluating measures of association (Blalock 1972).

Research Question #2. The second research question in phase I (i.e., which variable (or set) contributes the most toward the total mutual association) is addressed by computing the differences between the overall association, $T(A:B:C:D)$, and each of the conditional associations, $T_A(BCD)$, $T_B(ACD)$, $T_C(ABD)$, $T_D(ABC)$, $T_{AB}(CD)$, $T_{AC}(BD)$, $T_{AD}(BC)$, $T_{BC}(AD)$, $T_{BD}(AC)$, and $T_{CD}(AB)$. Equation (10) calculates the transmission difference ($\Phi_i T$):

$$(10) \quad \begin{aligned} \Phi_1 T &= T(A:B:C:D) - T_A(B:C:D) \\ \Phi_2 T &= T(A:B:C:D) - T_B(A:C:D) \\ \Phi_3 T &= T(A:B:C:D) - T_C(A:B:D) \\ \Phi_4 T &= T(A:B:C:D) - T_D(A:B:C) \\ \Phi_5 T &= T(A:B:C:D) - T_{AB}(C:D) \\ \Phi_6 T &= T(A:B:C:D) - T_{AC}(B:D) \\ &\text{etc.} \end{aligned}$$

The largest difference would identify the variable with the highest contribution to the mutual association (or equivalently, with the lowest remaining association among the unknown variables). The four-variable system in Figure 4 will be used to demonstrate questions 2, 3, and 4 in phase I, graphically:

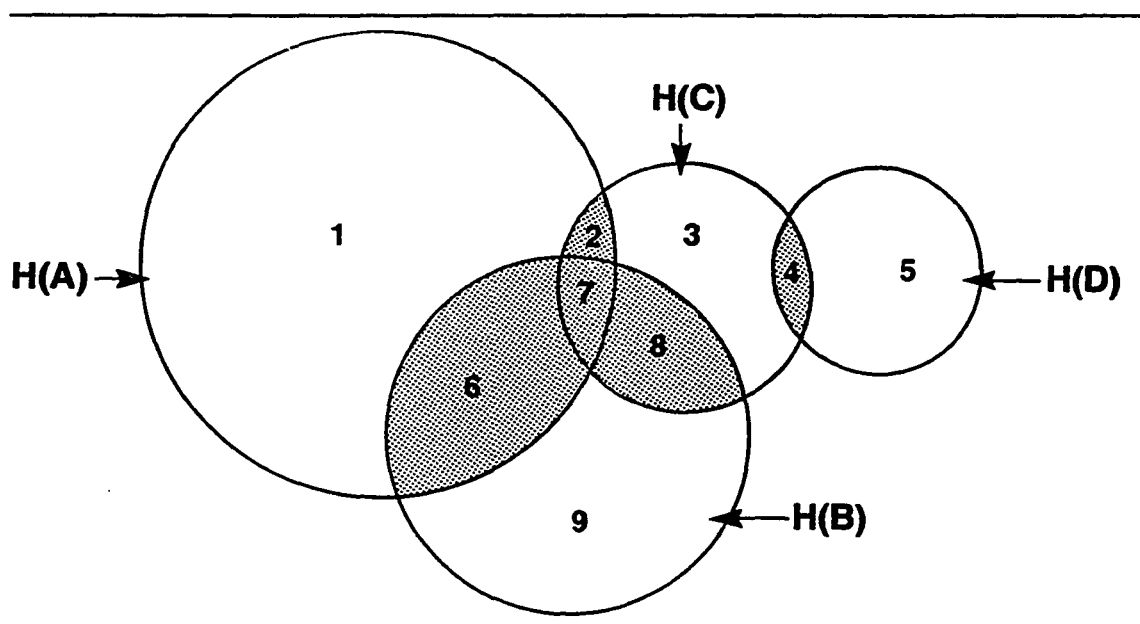


Figure 4. A System of four variables.

Variable B clearly includes most of the mutual association (as measured by transmission) among these four variables. The total mutual association is composed of areas 2, 4, 6, 2*7, and 8. Variable B accounts for 6, 7, and 8 which is the highest amount as compared to 2, 6, and 7

accounted for by A, 2, 7, and 8 accounted for by C, and 4 accounted for by D.

A significance test can be performed on any of the equations in (3). Here, the null hypothesis will be stated as, e.g., $T_A(B:C:D)$ is not consistent with the independence model for (B,C,D) subtables within each class of A (or, equivalently, $T(A:B:C:D) - T_A(B:C:D)$ fits the data). Then, small values of chi-square would reject the null hypothesis (or large values of the alternative model would reject its null hypothesis) and establish significance for the given variable. This is because if much of the mutual association is due to the known variable, then, holding it constant would render the rest of the variables independent (i.e., non-communicating, in information theoretic terms)

Alternately, partial-normalized transmissions or partial covariability coefficients can be used to measure the changes in the overlap given each variable. Here, then, a threshold criterion (e.g., 20%) would be established to judge these measures with.

Research Question #3. The previous research question addressed the contribution of each segmentation variable to the total mutual association among the segmentation variables. The third research question assesses this contribution to the overall joint variability among these variables.

To calculate the joint variability in a system Krippendorff's measure of systematic entropy will be used. Systematic entropy is computed by subtracting the noise factor for each variable from the total uncertainty in the system:

$$(11) S(W:X:Y:Z) = H(W, X, Y, Z) - H_{XYZ}(W) - H_{WYZ}(X) - H_{WXZ}(Y) - H_{WXY}(Z)$$

This calculation, as shown in Equation (11), in effect excludes the uncertainty in each variable which is not accounted for by any other variable (i.e., the noise). What remains is a "one-layer" measurement of the overlap in the system. Figure 5 depicts a four-variable system. The shaded area is the systematic entropy of this system.

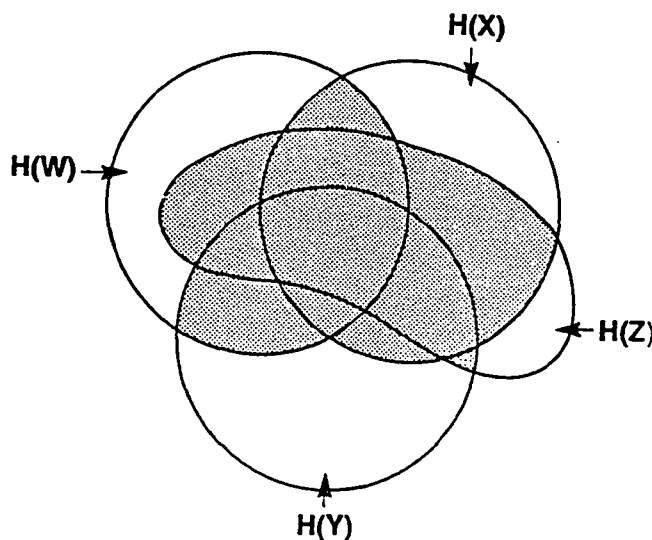


Figure 5. Systematic entropy of a four-variable system.

Partial systematic entropies can then be used to assess the contribution of each variable to the joint variability in the system. The information theoretic equations for the four-variable case developed below:

$$\begin{aligned}
 (12) \quad \Phi_1 \quad S &= S(A:B:C:D) - S_A(B:C:D) \\
 \Phi_2 \quad S &= S(A:B:C:D) - S_B(A:C:D) \\
 \Phi_3 \quad S &= S(A:B:C:D) - S_C(A:B:D) \\
 \Phi_4 \quad S &= S(A:B:C:D) - S_D(A:B:C) \\
 \Phi_5 \quad S &= S(A:B:C:D) - S_{AB}(B:C) \\
 \Phi_6 \quad S &= S(A:B:C:D) - S_{AC}(B:D) \\
 \text{etc.}
 \end{aligned}$$

Interestingly, the above equations are identical to difference between (n)-variable and (n-k)-variable transmissions, where, n refers to the first term in the above equations (e.g., A:B:C:D), and n-k refers to the unknown portion of the second term (e.g., B:C:D). The proof for this assertion follows:

The three variable case

$$\begin{aligned}
 S(ABC) - S_A(BC) &= H(ABC) - H_{BC}(A) - H_{AC}(B) - H_{AB}(C) \\
 &\quad - (H_A(BC) - H_{AC}(B) - H_{AB}(C)) \\
 &= H(ABC) - H_{BC}(A) - H_A(BC) \\
 &= H(ABC) - H(ABC) + H(BC) - H(ABC) + H(A) \\
 &= -H(ABC) + H(BC) + H(A) \\
 &= H(A) + H(B) + H(C) - H(ABC) - H(B) - H(C) + H(BC) \\
 &= [H(A) + H(B) + H(C) - H(ABC)] \\
 &\quad - [H(B) + H(C) - H(BC)] \\
 &= T(ABC) - T(AB)
 \end{aligned}$$

In General

$$\begin{aligned}
& S(ABC\dots) - S_A(BC\dots) \\
&= H(ABC\dots) - H_{BC\dots}(A) - H_{AC\dots}(B) - H_{AB\dots}(C) - \dots \\
&\quad - (H_A(BC\dots) - H_{AC\dots}(B) - H_{AB\dots}(C) - \dots) \\
&= H(ABC\dots) - H_{BC\dots}(A) - H_A(BC\dots) \\
&= H(ABC\dots) - H(ABC\dots) + H(BC\dots) - H(ABC\dots) + H(A) \\
&= -H(ABC\dots) + H(BC\dots) + H(A) \\
&= H(A) + H(B) + H(C) + \dots - H(ABC\dots) - H(B) - H(C) - \dots + H(BC\dots) \\
&= [H(A) + H(B) + H(C) + \dots - H(ABC\dots)] \\
&\quad - [H(B) + H(C) + \dots - H(BC\dots)] \\
&= T(ABC\dots) - T(BC\dots)
\end{aligned}$$

Therefore, a significance test can easily be performed (e.g., the null hypothesis is $T(A:B:C:D) - T(B:C:D)$ adequately represents the system, i.e., no significant difference between the expected frequencies generated through products of marginals and the observed values, resulting in a statistically insignificant (or small) likelihood-ratio chi-square value). Here, obviously, sufficiently large values of chi-square would result in the rejection of the null hypotheses.

Again, a normalized measure may be developed which would reveal the amount, rather than just the significance, of these measures. This measure is $1 - S_i/S$, where S is the total systematic entropy in the system and S_i is the systematic entropy remaining once variable i is known. This measure is equal to zero only if knowing i does not reduce

the systematic entropy, and it is equal to one if i accounts for all of the systematic entropy.

Research Question #4. The fourth research question in phase I (i.e., which variable(s) contributes the most toward the total uncertainty in the system of segmentation bases) is assessed using a similar approach as that used for the second question; however, here, instead of transmission differences, uncertainty differences are computed using equation (7). This is shown in the set of equations in (13) below:

$$\begin{aligned}
 (13) \quad & \Phi_1 H = H(A,B,C,D) - H(A) \\
 & \Phi_2 H = H(A,B,C,D) - H(B) \\
 & \Phi_3 H = H(A,B,C,D) - H(C) \\
 & \Phi_4 H = H(A,B,C,D) - H(D) \\
 & \Phi_5 H = H(A,B,C,D) - H(AB) \\
 & \Phi_6 H = H(A,B,C,D) - H(AC) \\
 & \text{etc.}
 \end{aligned}$$

In Figure 4, judging spatially, variable A accounts for most of the area of uncertainty occupied by the four-variable system. Therefore, knowledge of variable A would reduce the uncertainty about the total system by the most amount

Research Question #5. In order to assess the fifth research question in phase I (i.e., which variable has the greatest number of significant dyadic transmissions with the remaining variables) a series of two-way transmission tables are utilized in the following form:

(14) T(A:B), T(A:C), T(A:D)
 T(B:A), T(B:C), T(B:D)
 T(C:A), T(C:B), T(C,D)
 etc.

Then, a likelihood-ratio chi-square test of significance is performed for each transmission value in the table and non-significant transmissions are eliminated. Finally, the variable with the greatest number of surviving transmissions with other variables is identified as the candidate for the distinguished base.

Again, referring to Figure 4, variable C has the greatest number of dyadic relations within the system (i.e., it has interrelations with A, B, and D), While A and C have two such relations and D has only one. By this criterion, variable C, therefore, would be most central in the four-variable system.

Phase II: Selection and Analysis of the Distinguished Segmentation Variable. This phase first selects a distinguished segmentation variable (or set of variables), then, proceeds with analyzing it in detail. So far, variables A, B, C, and D have been used to exemplify the segmentation variables. In order to continue with this example, variables W, X, Y, and Z will be introduced to exemplify exogenous descriptor variables.

The actual analyses will first use the other segmentation variables as descriptor variables, and next the analyses are repeated with the exogenous variables. Here, however, only the exogenous variables are used for the discussion of the methodological issues, as the two sets of analyses are identical with respect to the methods used.

Research Question #1. In order to address the first research question in phase II (i.e., which segmentation basis is best determined once we have knowledge of a set of relevant descriptor variables?), the variability (entropy) in each candidate distinguished segmentation base, obtained in phase I, can be decomposed into the portion determined by the descriptor variables and the portion which remains unexplained. The expression, $H_{XYZ}(A) [=H(AXYZ)-H(XYZ)]$ is interpreted as variability in A unaccounted for by variables X, Y, and Z. The descriptor variables X, Y, Z, etc. can be selected either from the set of segmentation variables or other (exogenous) variables. Similar expressions are then obtained for other candidate segmentation variables:

$$\begin{aligned}
 (15) \quad & \Phi H(A) = H(A) - H_{XYZ}(A) \\
 & \Phi H(B) = H(B) - H_{XYZ}(B) \\
 & \Phi H(C) = H(C) - H_{XYZ}(C) \\
 & \Phi H(D) = H(D) - H_{XYZ}(D) \\
 & \Phi H(AB) = H(AB) - H_{XYZ}(AB) \\
 & \Phi H(AC) = H(AC) - H_{XYZ}(AC) \\
 & \text{etc.}
 \end{aligned}$$

Equation (15) computes the reduction in uncertainty of the segmentation variable due to the descriptor variables X , Y , and Z . The segmentation variable with the greatest reduction in its uncertainty is then said to be determined "best" by these descriptor variables. This variable can then be designated as the "distinguished" variable.

Figure 6 presents $H_{XYZ}(A)$ graphically. The shaded portion of the area of uncertainty for A is the portion represented by the (descriptor) variables X , Y , and Z .

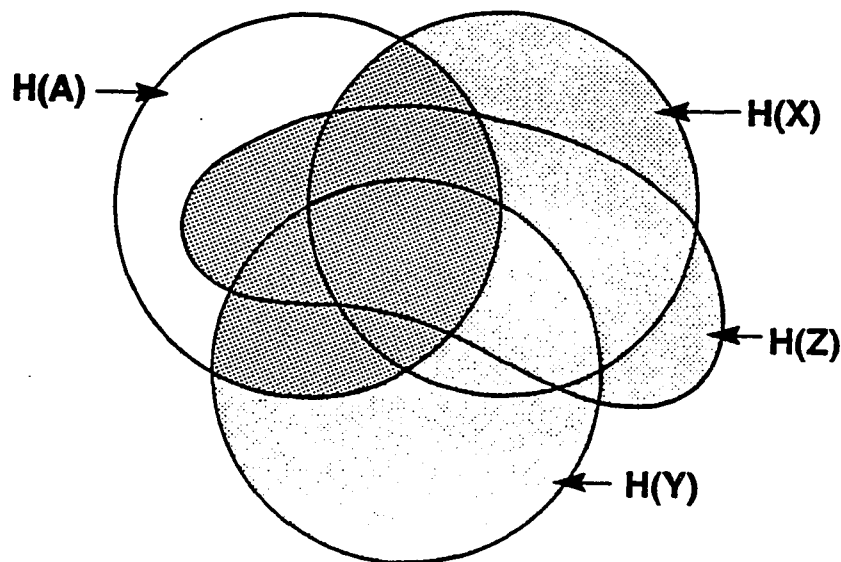


Figure 6. Determination of variable A with the knowledge of the descriptor variables X , Y , and Z .

As an example of how this method works, consider the case where the other segmentation variables are used as

descriptors. Therefore, if, for example, variable A is under investigation as a potential distinguished variable, B, C, and D are considered the descriptor variables. For instance, considering variables A, C, and D in Figure 4; it is evident that a greater proportion (or percent) of the uncertainty in C is accounted for with the knowledge of the other variables, followed by A and D. Visually, the following relationship between the proportions of uncertainties accounted for seems to exist for these three variables:

$$(16) \quad \frac{H(C) - H_{ABD}(C)}{H(C)} > \frac{H(A) - H_{BCD}(A)}{H(A)} > \frac{H(D) - H_{ABC}(D)}{H(D)}$$

or, as represented by the areas in figure 4:

$$(2,4,7,8)/(2,3,4,7,8) > (2,6,7)/(1,2,6,7) > 4/(4,5).$$

Research Question #2. This research question uses the algorithm developed for the first research question in this phase to assess the predictability (or reduction in uncertainty) of the distinguished segmentation variable under various collapsing schemes. The criteria for aggregation of classes (i.e., collapsing schemes) are subjective and based on the needs of the decision maker.

In terms of electricity consumption one may collapse the "very low," "low", and "medium," electricity consumption

groups as the first class, and high and very high groups as the second class. Here, the predictability of each of electricity consumption can be assessed in a dichotomized variable which is easy to interpret.

Research Question #3. The third question in Phase II (i.e., prioritizing the effect of the independent variables on the distinguished set) can be assessed using a method proposed by Ashby (1964) and used by Abrahamse and van Bueren (1980) and Krippendorff (1981). In this method, first marginal entropy (marginal variability) is considered for the variable of interest (e.g., A), then, bivariate transmissions are computed between the segmentation variable and the descriptor variables (i.e., $T(A:W)$, $T(A:X)$, $T(A:Y)$, $T(A:Z)$). These bivariate transmissions are in turn used to identify the descriptor variable that contributes the most to the segmentation variable. Then conditional bivariate transmissions are computed using the variable obtained above -- i.e., assuming X is selected as the variable with the highest bivariate transmission with A, $T_X(A:Y)$ and $T_X(A:Z)$ are computed, and the largest of these is selected. Repeating this process for all of the independent variables will result in some partitioning similar to the following:

$$(17) H(A) = T(A:X) + T_X(A:Y) + T_{XY}(A:Z) + H_{XYZ}(A)$$

where, X is the variable with the highest bivariate transmission with A ; Y is the variable with the highest bivariate transmission with A , given X ; Z is the variable with the highest bivariate transmission with A , given X and Y ; and $H_{XYZ}(A)$ is the remaining variability in A not explained by variables X , Y , or Z . Therefore, we can assert that: Of the variability in variable A , the amount equivalent to $T(A:X)$, $T_X(A:Y)$, and $T_{XY}(A:Z)$ are explained by variables X , Y , and Z , respectively, while $H_{XYZ}(A)$ is exclusive to variable A .

Figure 7 illustrates this situation. Variable A is assumed to be the segmentation variable. Again, judging spatially, variable Z seems to be the highest contributor as (4,5,11) occupies the largest area of uncertainty as compared to W (i.e., 3,4,8), X (i.e., 8,9) and Y (i.e., 10, 11). It is also apparent that W is the next highest contributor as (3,8) (i.e., $T_Z(A:W)$) is larger than (10) (i.e., $T_Z(A:Y)$) and $T_Z(A:Y) = (8,9)$. By the same token, Y and X are next in the ranking, respectively.

Research Question #4. The fourth question in phase II is stated as: What is the dependency structure in each class of the distinguished segmentation base? This question can be approached by utilizing the technique for "dependency analysis" suggested by Conant (1982).

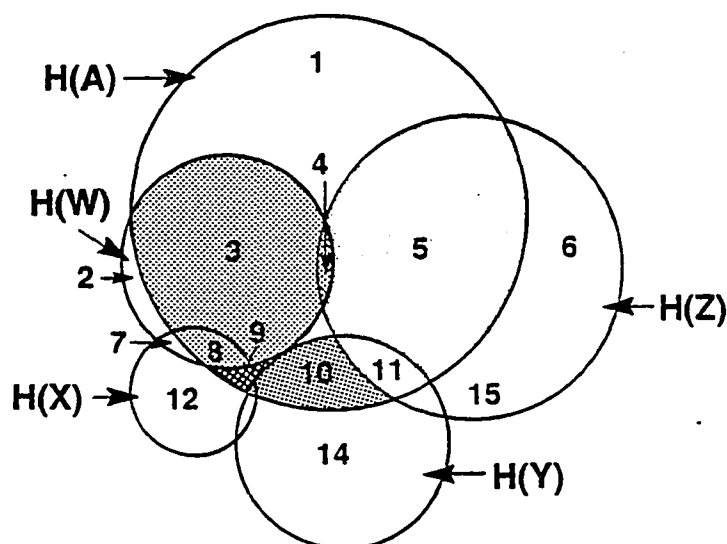


Figure 7. Prioritization of the contributions of variables W, X, Y, and Z to determine variable A.

In dependency analysis, first bivariate transmission between each variable (e.g., W) and the remaining variables in the system is computed. Then the single variable with the highest transmission is identified.--i.e., $\text{Max}\{T(W:X), T(W:Y), T(W:Z)\}$. Next, bivariate transmissions between the variable under consideration (i.e., W) and two-variable combinations, taken as a whole, of other variables (i.e., XY, XZ, YZ) are computed in order to determine the two-variable component that has the highest transmission with the variable under investigation --i.e., $\text{Max}\{T(W:XY), T(W:XZ), T(W:YZ)\}$. Next, bivariate transmissions between the variable under investigation (i.e., W) and three-

variable combinations, taken as a whole, of the remaining variables are computed such that the maximum is identified (note that, here, there is only one three-variable combination).

Next, another variable is considered (e.g., X). In the same manner as above, then, the highest bivariate transmissions with single, double, triple, etc. variables are identified. Subsequently, a table similar to Table VI is constructed.

In Table VI, Column 1 is the variable under consideration, Column 2 is the number of variables in the combination S_i , Column 3 is the S_i which has the highest bivariate transmission with the variable in Column 1, Column 4 lists bivariate transmissions between the variables in Column 1 and 3, the latter taken as an aggregate variable. Column 5 is column 4 divided by the entropy of the variable in Column 1, Column 6 is Column 5 as a percent of $\text{Max}(T/H(i))$, and Column 7 is the incremental effect of variables for each group measured by the transmission differences.

This table is in turn used to produce graphs of each variable and other variables showing the strength of relationships.

TABLE VI
CALCULATIONS FOR DEPENDENCY ANALYSIS OF WXYZ

(1)	(2)	(3)	(4)	(5)	(6)	(7)
i	n	S_i	$T(i:S_i)$	$T_n/H(i)$	%	$T(n)-T(n-1)$
W	1	X	$T(W:X)$	$T_1/H(W)$	*	$T(W:X)$
	2	XY	$T(W:XY)$	$T_2/H(W)$	*	$T(W:XY) - T(W:X)$
	3	XYZ	$T(W:XYZ)$	$T_3/H(W)$	*	$T(W:XYZ) - T(W:XY)$
X	1	W	$T(X:W)$	$T_1/H(X)$	*	$T(X:W)$
	2	WY	$T(X:WY)$	$T_2/H(X)$	*	$T(X:WY) - T(X:W)$
	3	WYZ	$T(X:WYZ)$	$T_3/H(X)$	*	$T(X:WYZ) - T(X:WY)$
Y	1	X	$T(Y:X)$	$T_1/H(Y)$	*	$T(Y:X)$
	2	XW	$T(Y:XW)$	$T_2/H(Y)$	*	$T(Y:XW) - T(Y:X)$
	3	XWZ	$T(Y:XWZ)$	$T_3/H(Y)$	*	$T(Y:XWZ) - T(Y:XW)$
Z	1	X	$T(Z:X)$	$T_1/H(Z)$	*	$T(Z:X)$
	2	XY	$T(Z:XY)$	$T_2/H(Z)$	*	$T(Z:XY) - T(Z:X)$
	3	XYW	$T(Z:XYW)$	$T_3/H(Z)$	*	$T(Z:XYW) - T(Z:XY)$

* $T(i:S_i)$ as a percent of the $\max(T(i:S_i))$ in the block i.

In order to construct such graphs, Conant (1981, 1982) suggests that a threshold level be selected based on which components of the dependency structure are identified. For instance, a threshold level of 80% would suggest that for each variable i (Column 1) in Table VI level n (Column 2) is selected such that the % (Column 6) is at most (to be explained shortly) 80%. The level n, as selected in the above manner would, then, determine what other variables are included in the component which includes variable i.

The above procedure would ensure that variables included in a component include at most 80% of the transmission between variable i and the remaining variables. Then, the variables suggested by level n for each i are used to form graphs of each component. For instance, let us assume that $n=2$ for $i=W$. This indicates that variables X and Y should be included in that component. Further, suppose that $T(n)-T(n-1)$ (Column 7) for the above X and Y are 0.4 and 0.1, respectively. This would suggest that the first component on the dependency structure be WXY , with a graph in which the width of the line indicating the relation between W and X , and W and Y , represented such that the contributions of the variables X and Y to the transmission $T(W:XY)$ are reflected.

In order to ensure that the sequence of inclusion of variables in each component (i.e., which variable was selected first, second, etc.) is reflected, they are numbered (as 1, 2, etc.) on the graph of each component. This will indicate that the contribution of a variable selected as number two assumes the inclusion of the first variable (i.e., it shows the incremental contribution to the variable i), and so on.

Finally, after all of the components are selected and their graphs are constructed, a combined graph is discerned from the individual graphs. It should be noted that, often, the width of the line between two variables determined by two different levels of i do not coincide (i.e., they do not show the same amount of contribution). The reason for this is the fact that the two width show incremental effects based on two separate conditions (i.e., preceding variables). In these cases, Conant selected the wider line, a convention which is followed here. Hypothetical graphs of the four components shown in Table VI and the final graph which emerges from them are illustrated in figure 8.

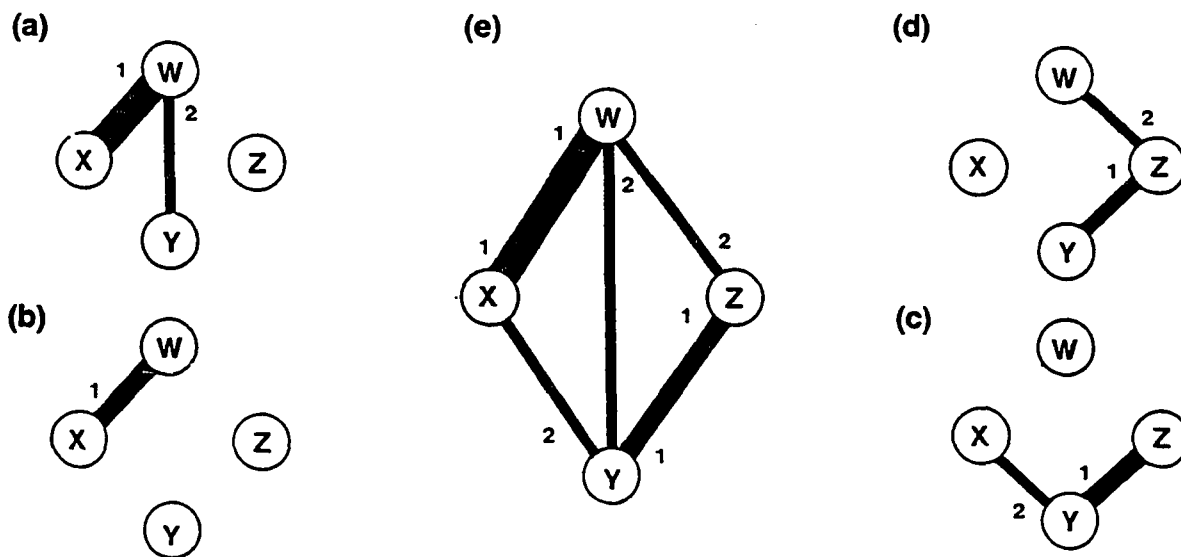


Figure 8. A sample dependency analysis diagram.

In order to discern the structural model (in the traditional sense, e.g., $WXY/XZ/YZ$) the following procedure is used. First, all components are included. Thus, based on Figure 8, $WXY/WX/WYZ/XYZ$ is obtained. Then, all of the embedded elements (i.e., components included in another component) as well as all but one of the identical elements are removed. This will produce $WXY/WXZ/XYZ$ in the above example. This will produce a structure of the data which is an approximation of the optimal structural model (Broekstra 1981, Conant 1981, 1982).

Computer Program

A computer program was developed to perform the information theoretic analyses. SYSENT is a FORTRAN program, available in both mainframe and microcomputer formats. It performs calculations of entropies, transmissions, and systematic entropies for an arbitrary number of variables (currently set at 10) and an arbitrary number of classes for each variable (currently set at 10).

SYSENT takes as input NDIM, the number of variables in the system, IDIM(i), $i=1,2,\dots,NDIM$, number of classes per variable, and LTABLE(j), $j=1,2,\dots,Prod(IDIM(i), i=1,2,\dots,NDIM)$, cell frequencies (or probabilities) for the contingency table, and produces as output, all marginal and conditional entropies, transmissions, systematic entropies. SYSENT also produces tables of values to answer a number of

specific research questions included in this study. Appendix V includes a samples of the output file for SYSENT.

The General Log-Linear/Logit Approach

This section develops the general log-linear and logit expressions and models used to address the research questions, where applicable. As will be demonstrated, certain research questions have been addressed by Green and Carmone or others, in the development of segment congruence analysis while certain other questions have not applied in that context.

Of those questions which have not been addressed in segment congruence analysis, certain ones have been addressed in other contexts, using log-linear models, while certain others, though not addressed by any major published material, could conceivably be addressed by these models. Yet a fourth category of research questions, or certain aspects within them, cannot be addressed by the general log-linear models. The reader unfamiliar with the general log-linear and logit models should refer to Appendix I for a brief introduction to these models.

This section employs the same variables which were used to discuss the information theoretic approach. I.e., the segmentation variables are assumed to be A, B, C, and D;

while the descriptor variables are assumed to be X, Y, and Z.

Phase I: Analysis of The Segmentation Variables. As noted in Chapter 2, one major shortcoming of the general log-linear models is their inability to generate a measure of the amount of association. In addressing the first three questions, the information theoretic approach generated methods of measuring the amount of association as well as a measure of the test of independence and/or significance. In applying the general log-linear models, however, only a measure of significance (of approximation to the system by the model) is generated. (One should note, however, that significance tests for both information theoretic and log linear methods are actually done using the likelihood-ratio chi-square).

Research Question #1. The general log-linear model for independence (i.e., the main-effects model is used to address research question #1 (i.e., Are the segmentation variables mutually associated? If yes, how can this mutual association be measured?). This model for a four-variable system ABCD has the following form:

$$(18) \quad \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D$$

If the null hypothesis of independence is rejected for the above model, it is concluded that statistically

significant overlap exists among the variables in the multivariate system. The likelihood-ratio chi-square is used to test the null hypothesis.

No measure of the amount of overlap, which is an important issue in segment congruence analysis, is provided by this approach.

Research Question #2. The second research question (i.e., Which basis (or set of bases) makes the highest contribution toward the mutual association?) can be addressed using the log-linear model, for all two-way interaction effects. Both Green and Carmone (1977) and Van Auken and Lonial (1984) have misinterpreted this question. They assessed the contribution of each segmentation variable to the joint variability (systematic entropy) in the system, but used it for assessing mutual association (transmission). It is given (Broekstra 1982) that:

$$(19) T_A(B:C:D) = T(A:B:C:D) - T(A:B) - T(A:C) - T(A:D)$$

$T(A:B:C:D) - T_A(B:C:D)$ assesses the contribution of A to the mutual association (i.e., the effect of A being constant, or given, and this, as just shown, is equal to the sum of the bivariate transmissions including A). The single-variable independence model in the log-linear approach, which is presented later in this chapter, examines

contributions not to mutual association but to the joint uncertainty as follows:

$$(20) \quad T(A:B:C:D) - T(B:C:D), \text{ or equivalently;} \\ S(A:B:C:D) - S_A(B:C:D).$$

This was established by the fact that the same likelihood-ratio chi-square obtained for the above information theoretic model and the single-variable independent model, using both Green and Carmone data and the data analyzed in this study.

To assess the contribution of each variable (or set of variables) to the total mutual association, models similar to the following must be employed:

$$(21) \quad \begin{aligned} \text{LL: } (A:B)(A:C)(A:D) &\Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C \\ &\quad + u_l^D + u_{ij}^{AB} + u_{ik}^{AC} + u_{il}^{AD} \\ (B:A)(B:C)(B:D) &\Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C \\ &\quad + u_l^D + u_{ij}^{AB} + u_{jk}^{BC} + u_{jl}^{BD} \\ (C:A)(C:B)(C:D) &\Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C \\ &\quad + u_l^D + u_{ik}^{AC} + u_{jk}^{BC} + u_{kl}^{CD} \\ (D:A)(D:B)(D:C) &\Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C \\ &\quad + u_l^D + u_{il}^{AD} + u_{jld}^{BD} + u_{kl}^{CD} \end{aligned}$$

This model can then be compared with the independence model with respect to the drop in its likelihood ratio chi-square. The null hypotheses are stated as; e.g., the model $\{AB\}\{AC\}\{AD\}$ fits the data. If this were true, a small

likelihood ratio chi-square would be obtained, which, in turn, would result in tentatively accepting the hypothesis that A is significantly associated with B, C, and D.

In order to understand the implications of the above model, one can use its equivalent test in information theory. As mentioned, $\{AB\}\{AC\}\{AD\}$ is equivalent to $T_A(BCD)$. Testing the above log-linear model is, therefore, equivalent to testing whether, given A; B, C, and D are independent (i.e., $H_0: T_A(BCD)=0$). Accepting such hypothesis (which will result from small likelihood-ratio chi-square) would mean that the significant association observed in the first research question is due to A and not B, C, or D. Rejection of such hypothesis would indicate that, at least, by holding A constant, there still exists association among the remaining variables. Here, then, one would search for the variable which removes a greater amount of association than the rest.

For this model, also, the log-linear approach does not provide a measure of strength of the relation. The log-linear analysis only indicates whether a model fits statistically, but not how much overlap exists.

Research Question #3. In order to address the third research question (i.e.,_which basis, or set of bases, makes the highest contribution toward the joint variability in the system of segmentation bases?), the single-variable (or single-clustering) log-linear model is utilized:

$$\begin{aligned}
(22) \quad \{A\}\{BCD\} &\Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D \\
&\quad + u_{jk}^{BC} + u_{jl}^{BD} + u_{kl}^{CD} + u_{jkl}^{BCD} \\
\{B\}\{ACD\} &\Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D \\
&\quad + u_{ik}^{AC} + u_{il}^{AD} + u_{kl}^{CD} + u_{ikl}^{ACD} \\
\{C\}\{ABD\} &\Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D \\
&\quad + u_{ij}^{AB} + u_{il}^{AD} + u_{jl}^{BD} + u_{ijl}^{ABD} \\
\{D\}\{ABC\} &\Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D \\
&\quad + u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC} + u_{ijk}^{ABC} \\
\{AB\}\{CD\} &\Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D \\
&\quad + u_{ij}^{AB} + u_{kl}^{CD} \\
\{AC\}\{BD\} &\Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D \\
&\quad + u_{ik}^{AC} + u_{jl}^{BD}
\end{aligned}$$

This model hypothesizes that the model for the independence of the segmentation variable (or set) under the consideration is a good model of the data. Likelihood-ratio chi-square is used here to assess this hypothesis for each variable of group of variables. Small values for the likelihood ratio chi-squares would indicate that its corresponding model (e.g., $\{A\}\{BCD\}$) fits the data adequately, and, therefore the variable under consideration is independent of the rest (e.g., A is independent of BCD).

Again, the log-linear approach, unlike the information theoretic approach, does not generate a direct measure of association among the variables in any sense. It should be mentioned, however, that the likelihood-ratio chi-squares generated by the models for research questions one thru

three have been used as measures of association as well. Both Green and Carmone (1977) and Van Auken and Lonial (1984) used the likelihood-ratio chi-square in this way.

Research Question #4. Research question #4 (i.e., Knowledge of which base (or set of bases) makes the highest contribution toward the reduction of uncertainty in the overall system of segmentation bases?) has not been addressed in the context of log-linear models. This is due to the fact that the notion of total uncertainty in a system has not been investigated in this area. However, it is theoretically possible to address this issue by examining the likelihood-ratio chi-square for the no-effects model. The measure obtained is a linear function of the information theoretic measure of uncertainty (i.e., entropy):

The no-effects is stated as:

$$(23) \quad \log \hat{F}_{ij} = u$$

where

$$u = (1/IJK...) \sum_i \sum_j \sum_k \dots \log F_{ijk} \dots$$

$$\begin{aligned} \text{then } L_{LL}^2 &= -2N \log(2)(H) + 2N \log(A) \\ &= -2N \log(2)[H - H_{MAX}] \end{aligned}$$

Obviously, calculating H alone is much simpler (and more logical) than using the log-linear no-effect model and then testing it for significance using the likelihood ratio chi-square. This becomes particularly evident if it is

necessary to assess the difference between two uncertainties (as is the case here).

In order to assess the contribution of a variable to the total uncertainty in the system, using the general log-linear model, the no-effects model for both the total system and the variable under consideration must be examined. Then, the uncertainty should be computed by linear transformations of the likelihood-ratio chi-square for both methods (i.e., such that the H terms are calculated). Then, the difference between the two (i.e., the overall system and the variable under the consideration) should be computed.

Furthermore; most standard computer packages (e.g., SAS or SPSS^X) do not even allow the consideration of the no-effects model. It would be very cumbersome to perform the necessary calculations manually particularly for large tables. This is due to the requirement by the log-linear models to generate the expected cell frequencies by the maximum likelihood estimation method first and, then, calculate the likelihood-ratio chi-square.

Research Question #5. The fifth research question (i.e., Which variable (or set of variables) has the greatest number of significant interrelationships with other variables?) is addressed by examining the log-linear independence models two-variable combinations of the system of segmentation variables.

$$\begin{aligned}
 (24) \quad \{A\}\{B\} &\Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B \\
 \{A\}\{C\} &\Rightarrow \log P_{ijkl} = u + u_i^A + u_k^C \\
 \{A\}\{D\} &\Rightarrow \log P_{ijkl} = u + u_i^A + u_l^D \\
 \{B\}\{A\} &\Rightarrow \log P_{ijkl} = u + u_j^B + u_i^A \\
 \{B\}\{C\} &\Rightarrow \log P_{ijkl} = u + u_j^B + u_k^C \\
 \{B\}\{D\} &\Rightarrow \log P_{ijkl} = u + u_j^B + u_l^D \\
 &\text{etc.}
 \end{aligned}$$

The variable for which more of these bivariate models are significant (i.e., greatest number of bivariate combinations for it depart from independence significantly) is regarded as the variable with the greatest number of dyadic relations. This variable, therefore, plays a more central role as it has interactions with more components in the system.

Here, also, the information theoretic approach provides a direct measure of association, while, the log-linear approach does not.

Phase II: Selection and Analysis of the Distinguished Segmentation Variable. Using the log-linear/logit approach the four research questions in this phase are approached as follows:

Research Question #1. Which one of the candidate distinguished segmentation bases, identified in phase I, is best predicted once we have the knowledge of a set of relevant independent variables (i.e., which set comprises the best distinguished segmentation base)? Multinomial

logit models are suitable in addressing this question. The simple (i.e., the two-by-two model) is presented below:

$$(25) \quad (A) \quad BY(X)(Y)(Z) \Rightarrow O_{jkl}^A = b^A + b_j^{AX} + b_k^{AY} + b_l^{AZ} + e$$

$$(B) \quad BY(X)(Y)(Z) \Rightarrow O_{jkl}^B = b^B + b_j^{BX} + b_k^{BY} + b_l^{BZ} + e$$

$$(C) \quad BY(X)(Y)(Z) \Rightarrow O_{ijkl}^C = b^C + b_j^{CX} + b_k^{CY} + b_l^{CZ} + e$$

$$(D) \quad BY(X)(Y)(Z) \Rightarrow O_{ijkl}^D = b^D + b_j^{DX} + b_k^{DY} + b_l^{DZ} + e$$

where; e.g.,

$$O_{ijkl}^A = \log(P_{ijkl}^{A=1} / 1 - P_{ijkl}^{A=1}),$$

b^A , b_j^{AX} , b_k^{AY} , and b_l^{AZ} are the parameters estimated, and

e is the residual

All classes of each variable in the contingency table can be estimated as a function of the other variables in the model. The likelihood ratio chi-square is used to evaluate the model. The smaller the likelihood-ratio chi-square, the better the model fits the data.

The logit models shown above are known as the main-effects models. In general, if these models do not produce good estimations of cell frequencies statistically, higher-effects models will be introduced.

Research Question #2. Similar to the information theoretic counterpart of this research question, the logit models developed in the first research question can be utilized to address this research question. Here, the dependent variables are dichotomous, each one having a group

of particular interest as one class and the other groups as the second class.

Research Question #3. Green and Carmone (1977) briefly mentioned that the magnitude of the coefficients of the logit model can be used to address the third research question in phase II. This research question is stated as: How can one prioritize the effect of the independent variables on the distinguished set?

This approach is not practical, because the number of coefficients generated for each independent variable is equivalent to the degrees of freedom for a two way contingency table, with the dependent and the independent variable as its dimensions. In estimating each cell frequency from the parameters produced by the logit model, one need to decide which coefficients are present and then sum these coefficients. Then the antilogarithm of this sum gives the estimated odds of a particular class of the dependent variable against all other classes.

In the applications used by Green and Carmone (1977) and Van Auken and Lonial (1984), all of the variables were dichotomized. In such case, only one coefficient is generated for each two dimensional contingency table of the dependent variable and a dependent variable. Here, the coefficients can be used to judge the significance of each variable. However, there is no collective way to judge the significance of a variable based on its coefficients, as

suggested by Green and Carmone (1977), if any of the variables used is multichotomous.

The general log-linear model can be used to develop successive models (much the same as the information theoretic approach for this question) of significance of variables in a marginal and partial sense. First, a set of two-variable independence models (of the distinguished variable and the descriptor variables) are run and the most significant variable is identified (e.g., $\{A\}\{B\}$, $\{A\}\{C\}$, etc.). Then, a set of partial single-variable independence models are run composed of the distinguished variable versus the most significant variable found above and the remaining variables (e.g., assuming that B was selected as the most significant variable, $\{A\}\{CB\}$, $\{A\}\{CD\}$, etc.). Repeating this process to include all combinations (with the last one being $\{A\}\{BCDE...\}$) will result in a ranking identical to that of the information theoretic approach.

Research Question #4. The fourth research question (i.e., What is the dependency structure in each class of the distinguished segmentation base? How can it be quantified?) has neither been addressed directly, nor indirectly, in the segment congruence analysis context. Nevertheless, the HILOGLINEAR procedure in SPSSX, generated parameter estimates, which in turn can be used to develop a diagram of effects, including magnitudes of effects. This is the essence of dependency analysis.

Tables VII and VIII present the algebraic expressions, if applicable, for the research questions in the two phases, respectively.

Summary of the Methodology

Properties and theoretical underpinnings of information theory were detailed in the last section of Chapter 2 and in Appendix II. It was established that the information theoretic approach has numerous properties that make it a viable technique for data analysis. For instance, the decomposability property of informational measures and the analogy between these measures and variance (Garner and McGill, 1956) enables one to perform ANOVA like analyses of data with discrete dependent variables. Furthermore, this property provides a powerful structural modeling tool for researchers concerned with the analysis of nominally scaled data, in all fields (Broekstra, 1978, 1981, 1982, Krippendorff, 1979, 1981, Klir, 1976, 1986, etc.).

The methodology section of the present chapter established the fact that the information theoretic approach is capable of addressing all of the question concerning segment congruence analysis raised by this undertaking. Table IX presents a list of these research questions and the capability of the information theoretic approach as well as the log-linear approach to segment congruence analysis.

As Table IX suggests the information theory is capable of addressing all issues discussed within the framework of segment congruence analysis.

TABLE VII

INFORMATION THEORETIC AND LOG-LINEAR MATHE-
MATICAL EXPRESSIONS USED FOR SEGMENT
CONGRUENCE ANALYSIS (PHASE I)

Q I.1) Overlap

IT: $T(A:B:C:D) = H(A) + H(B) + H(C) + H(D) - H(A, B, C, D)$
 where;
 $H(A, B, C, D) = -\sum_i \sum_j \sum_k \sum_l P(A_i, B_j, C_k, D_l) \log_2 P(A_i, B_j, C_k, D_l)$

LL: $\{A\}\{B\}\{C\}\{D\} \Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D$

Q I.2) Contribution to Overlap

IT: $\begin{array}{ll} \Phi_1 & T = T(A:B:C:D) - T_A(B:C:D) \\ \Phi_2 & T = T(A:B:C:D) - T_B(A:C:D) \\ \Phi_3 & T = T(A:B:C:D) - T_C(A:B:D) \\ \Phi_4 & T = T(A:B:C:D) - T_D(A:B:C) \\ \Phi_5 & T = T(A:B:C:D) - T_{AB}(C:D) \\ \Phi_6 & T = T(A:B:C:D) - T_{AC}(B:D) \end{array}$
 etc.

LL: $\{A:B\}\{A:C\}\{A:D\} \Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D + u_{ij}^{AB} + u_{ik}^{AC} + u_{il}^{AD}$

$\{B:A\}\{B:C\}\{B:D\} \Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D + u_{ij}^{AB} + u_{jk}^{BC} + u_{jl}^{BD}$

$\{C:A\}\{C:B\}\{C:D\} \Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D + u_{ik}^{AC} + u_{jk}^{BC} + u_{kl}^{CD}$

$\{D:A\}\{D:B\}\{D:C\} \Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D + u_{il}^{AD} + u_{jl}^{BD} + u_{kl}^{CD}$

Table VII - (Continued)

Q I.3) Contribution to Joint Variability

IT: $\Phi_1 S = S(A:B:C:D) - S_A(B:C:D)$
 $\Phi_2 S = S(A:B:C:D) - S_B(A:C:D)$
 $\Phi_3 S = S(A:B:C:D) - S_C(A:B:D)$
 $\Phi_4 S = S(A:B:C:D) - S_D(A:B:C)$
 $\Phi_5 S = S(A:B:C:D) - S_{AB}(B:C)$
 $\Phi_6 S = S(A:B:C:D) - S_{AC}(B:D)$
 etc.

LL: $\{A\}\{BCD\} \Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D$
 $+ u_{jk}^{BC} + u_{jl}^{BD} + u_{kl}^{CD} + u_{jkl}^{BCD}$

$\{B\}\{ACD\} \Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D$
 $+ u_{ik}^{AC} + u_{il}^{AD} + u_{kl}^{CD} + u_{ikl}^{ACD}$

$\{C\}\{ABD\} \Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D$
 $+ u_{ij}^{AB} + u_{il}^{AD} + u_{jl}^{BD} + u_{ijl}^{ABD}$

$\{D\}\{ABC\} \Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D$
 $+ u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC} + u_{ijk}^{ABC}$

$\{AB\}\{CD\} \Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D$
 $+ u_{ij}^{AB} + u_{kl}^{CD}$

$\{AC\}\{BD\} \Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B + u_k^C + u_l^D$
 $+ u_{ik}^{AC} + u_{jl}^{BD}$

etc.

Table VII - (Continued)

Q I.4) Reduction of Uncertainty in the System

IT: $\Phi_1 H = H(A,B,C,D) - H(A)$
 $\Phi_2 H = H(A,B,C,D) - H(B)$
 $\Phi_3 H = H(A,B,C,D) - H(C)$
 $\Phi_4 H = H(A,B,C,D) - H(D)$
 $\Phi_5 H = H(A,B,C,D) - H(AB)$
 $\Phi_6 H = H(A,B,C,D) - H(AC)$
 etc.

LL: $\{.\}\{.\} \Rightarrow \log P_{ijkl} = u - \log P_i = u_i$
 $\{.\}\{.\} \Rightarrow \log P_{ijkl} = u - \log P_j = u_j$
 $\{.\}\{.\} \Rightarrow \log P_{ijkl} = u - \log P_k = u_k$
 $\{.\}\{.\} \Rightarrow \log P_{ijkl} = u - \log P_l = u_l$
 $\{.\}\{.\} \Rightarrow \log P_{ijkl} = u - \log P_i = u_{ij}$
 $\{.\}\{.\} \Rightarrow \log P_{ijkl} = u - \log P_i = u_{ik}$
 etc.

Q I.5) Number of Dyadic Relations

IT: $T(A:B), T(A:C), T(A:D)$
 $T(B:A), T(B:C), T(B:D)$
 $T(C:A), T(C:B), T(C,D)$
 etc.

LL: $\{A\}\{B\} \Rightarrow \log P_{ijkl} = u + u_i^A + u_j^B$
 $\{A\}\{C\} \Rightarrow \log P_{ijkl} = u + u_i^A + u_j^C$
 $\{A\}\{D\} \Rightarrow \log P_{ijkl} = u + u_i^A + u_l^D$

 $\{B\}\{A\} \Rightarrow \log P_{ijkl} = u + u_j^B + u_i^A$
 $\{B\}\{C\} \Rightarrow \log P_{ijkl} = u + u_j^B + u_j^C$
 $\{B\}\{D\} \Rightarrow \log P_{ijkl} = u + u_j^B + u_l^D$
 etc.

TABLE VIII

INFORMATION THEORETIC AND LOG-LINEAR MATHE-
MATICAL EXPRESSIONS USED FOR SEGMENT
CONGRUENCE ANALYSIS (PHASE II)

Q II.1) Predictability

IT: $\Phi H(A) = H(A) - H_{XYZ}(A)$
 $\Phi H(B) = H(B) - H_{XYZ}(B)$
 $\Phi H(C) = H(C) - H_{XYZ}(C)$
 $\Phi H(D) = H(D) - H_{XYZ}(D)$
 $\Phi H(AB) = H(AB) - H_{XYZ}(AB)$
 $\Phi H(AC) = H(AC) - H_{XYZ}(AC)$
 etc.

LL: {A} BY {X}{Y}{Z} $\Rightarrow O_{ijkl}^A = b^A + b_j^{AX} + b_k^{AY}$
 $+ b_l^{AZ} + e$

{B} BY {X}{Y}{Z} $\Rightarrow O_{ijkl}^B = b^B + b_j^{BX} + b_k^{BY}$
 $+ b_l^{BZ} + e$

{C} BY {X}{Y}{Z} $\Rightarrow O_{ijkl}^C = b^C + b_j^{CX} + b_k^{CY}$
 $+ b_l^{CZ} + e$

{D} BY {X}{Y}{Z} $\Rightarrow O_{ijkl}^D = b^D + b_j^{DX} + b_k^{DY}$
 $+ b_l^{DZ} + e$

where; e.g.,

$O_{ijkl}^A = \log(p_{ijkl}^{A=1} / 1 - p_{ijkl}^{A=1})$,
 b^A , b_j^{AX} , b_k^{AY} , and b_l^{AZ} are the parameters
 estimated, and
 e is the residual

Note: The above logit models are known as the main-effects models. In general, if these models do not produce good estimations of cell frequencies statistically, higher-effects models will be introduced.

Table VIII - (Continued)

Q II.2) Information/Noise ratio (Systematic entropy)

Same as Q II.1 after reclassifications of the distinguished variables.

Q II.3) Prioritization

IT: $H(A) = T(A:X) + T_X(A:Y) + T_{XY}(A:Z) + H_{XYZ}(A)$

LL: Magnitudes of coefficients in the logit model
(i.e., prioritize b_j^{AX} , b_k^{AY} , and b_l^{AZ})

Q II.4) Dependency Analysis

IT: Conant's (1982) Dependency Analysis

LL: The prioritization scheme developed in Q II.3

TABLE IX
COMPARISON OF THE TWO APPROACHES TO SEGMENT
CONGRUENCE ANALYSIS

NUMBER	QUESTION DESCRIPTION	INFO. THEORY	LOG-LINEAR	APPLIED TO SCA BEFORE
<u>Q I.1</u>	OVERLAP:			
	(EXISTENCE)	*	+	YES
	(AMOUNT)	*	#	NO
<u>Q I.2</u>	CONTRIBUTION TO OVERLAP:			
	(EXISTENCE)	*	#	NO
	(AMOUNT)	*	-	NO
<u>Q I.3</u>	CONTRIBUTION TO JOINT VAR:			
	(EXISTENCE)	*	+ ¹	YES
	(AMOUNT)	*	+ ¹	NO
<u>Q I.4</u>	CONTRIBUTION TO UNCERTAINTY:			
	(EXISTENCE)	*	^2	NO
	(AMOUNT)	*	^2	NO
<u>Q I.5</u>	NUMBER OF DYADIC RELATIONS	*	#	NO
<u>Q II.1</u>	PREDICTABILITY	*	+	YES
<u>Q II.2</u>	REFINEMENT OF PREDICTABILITY	*	#	NO
<u>Q II.3</u>	PRIORITIZATION	*	# ³	YES
<u>Q II.4</u>	DEPENDENCY ANALYSIS	*	-	NO

* = First done in this thesis (new programs)

= First done in this thesis (using SPSS^X)

+ = Reported earlier in literature

- = Has not been, but can easily be done

^ = Has not been done (too cumbersome w/o computer program)

- 1 The chi-square statistic has been used to make comparisons between models.
- 2 Appropriate no-effects models must be developed and their corresponding likelihood-ratio chi-squares must be computed to assess existence and amount.
- 3 Using coefficients of the logit function has been suggested for this purpose, but it only applies to cases where all variables are dichotomous.

THE DATA AND PROCEDURES

A survey conducted by the Bonneville Power Administration (BPA) in collaboration with the Pacific Northwest Utilities Conference Committee (PNUCC) and the Northwest Regional Power Planning Council (NRPPC) provides the basis for an empirical application of segment congruence analysis using the information theoretic approach. Among the existing techniques, the general log-linear and logit models are applied to this data as well, in order to demonstrate the efficacy of the information theoretic methods.

The Pacific Northwest Residential Energy Survey (PNWRES), was designed to gain knowledge on issues of energy use habits, attitudes towards energy use and conservation, and the extent of conservation activities (practices and installations) in the Pacific Northwest region. The information is used to support activities which are authorized by the Pacific Northwest Electric Power Planning and Conservation Act of 1980.

Description of the Study

The survey was administered during the summer months of 1983. The population was approximated at three million dwellings in the Pacific Northwest region. Of the original sample size of 6227 accounts, 546 were determined to be

ineligible, and 481 respondents refused to complete the interview. Also, interviewers were unable to contact an eligible respondent in 345 cases.

Overall, approximately nine percent (9%) of the accounts were determined to be ineligible and approximately eighty six percent (86%) of the eligible residents responded to the survey, an overall completion rate of over seventy six percent (76%). This amounted to a total sample size of 4703. The field work was conducted by Louis B. Harris and Associates, Inc.

According to the PNWRES User's Guide, "data collection began in late May 1983, and was completed in September 1983. Over 80 percent of the interviews were conducted in June and July." (PNWRES User's Guide, 1983) Eligibility of a respondent household consisted of existing service from an electricity utility to the home in question. Emphasis was placed upon residential meters, one residential unit per meter, and year-round as opposed to seasonal occupation of the home.

Average interview time was approximately one hour, and all surveys were edited, keypunched, and 100 percent verified. Ten percent of interviews were verified by phone.

First, utilities serving fewer than 1000 customers were grouped together or combined with larger utilities. Second, three of the larger privately owned utilities were disaggregated on the basis of political (state) boundaries

and geographic zones. These modifications to the clustering of utilities resulted in: Pacific Power and Light being subdivided and represented in every geographic and weather zone except Western Washington; the Washington Water Power Company was represented both in Eastern Washington and Idaho; and Idaho Power Company was represented both in Idaho and Eastern Oregon. This resulted in the generation of 116 mutually exclusive and exhaustive utility groupings.

A multistage cluster sampling technique involving two dimensions of stratification (by six levels of geographic regions and two levels of utility ownership, i.e., public and private) was employed for selecting the sample of utilities for 1983 PNWRES.

Within each of the six geographic strata composed of publicly owned utilities, substrata were constructed independently for each stratum, based upon the number of residential accounts served by utilities within that stratum. Individual utility clusters were then selected within substrata with probabilities proportionate to size, with the number of residential accounts as the measure of size.

All privately owned utilities within the sampling frame were selected with certainty into the survey sample. One pair of privately owned utilities was included in the sample as a utility cluster: Pacific Power and Light's Montana operations and Montana Light and Power Company.

Following selection of the 57 utility clusters, the number of subclusters (Meter Reading Routes or MRR) within a utility on which interviews were to be attempted was determined to ensure that substrata contributed relatively equal numbers of observations within strata and that strata contributed relatively equal numbers of observations to the total survey sample. Specific MRR's were selected via the proportional to size procedures. Finally, fifteen residential accounts were selected systematically by utilities for each MRR assigned.

Variables Measured. Nine general topics were included in the questionnaire for field interviews:

1. Basic dwelling unit characteristics.
2. Energy-related attitudes/opinions.
3. Conservation measures utilized.
4. Space heating fuels and equipment.
5. Water heating fuels and equipment.
6. Air conditioning fuels and equipment.
7. Household appliance characteristics.
8. Resident demographics.
9. Physical measurement of dwelling unit and water temperature.

With the consent of the respondent, interviewers measured the temperature of the hot water and the outside dimensions of the dwelling unit living space. Also, respondents were asked to sign a "Waiver Form" permitting

Bonneville Power Administration to access the respondents' electricity and, where appropriate, natural gas billing histories from relevant suppliers.

The billing history data includes both the amount of fuel used and the total cost of the fuel for each billing period between September 1981 and December 1982. In order to support the evaluation of Bonneville's residential weatherization program, thirteen utility clusters (nine different utilities) were asked to supply additional billing data for the period ending April 1983.

Additionally, data on weather and climatic changes were collected, in terms of daily maximum and minimum and heating-degree-days and cooling-degree-days for the period between September 1, 1981 and March 31, 1983. This data was collected with the help of National Oceanographic and Atmospheric Administration (NOAA).

Selection and Aggregation of Variables

A flow chart of procedures used for selection and aggregation of research variables is presented in Figure 9.

The variable selection process began by including all the variables which would (even remotely) lend themselves to a marketing analysis study in the PNWRES data. The arrows by the variable names in Appendix IV point to such variables. Six clusters of segmentation variables were identified using common sense and expert opinion. Several

officials from the Bonneville Power Administration, BPA, and the Pacific Power and Light Company, PP&L were interviewed and their judgments were taken into account in determination of relevant groups of variables.

Table X includes a listing of the variables used in the clusterings, their description and their corresponding question numbers in the questionnaire (see Appendix III for the questionnaire and Appendix IV for the complete variable listing).

SPSSX's clustering procedures were utilized to combine related variables into single composite variables. Table XI presents the list of these clusters, their description and the variable names, corresponding to the variable list in Appendix IV, for each cluster.

The questions related to four groups of these variables were so closely related or interwoven within each group that it was easily justified to cluster them together and treat them as aggregates of variables. These four groups included the energy related attitude questions, VAR049 thru VAR055, feelings and perceptions questions about the environment, VAR066 thru VAR073, awareness of funding for energy efficiency improvement programs, VAR074, VAR076, and VAR078, and major uses of electricity variables, VAR365, VAR367, VAR369, VAR371, and VAR373.

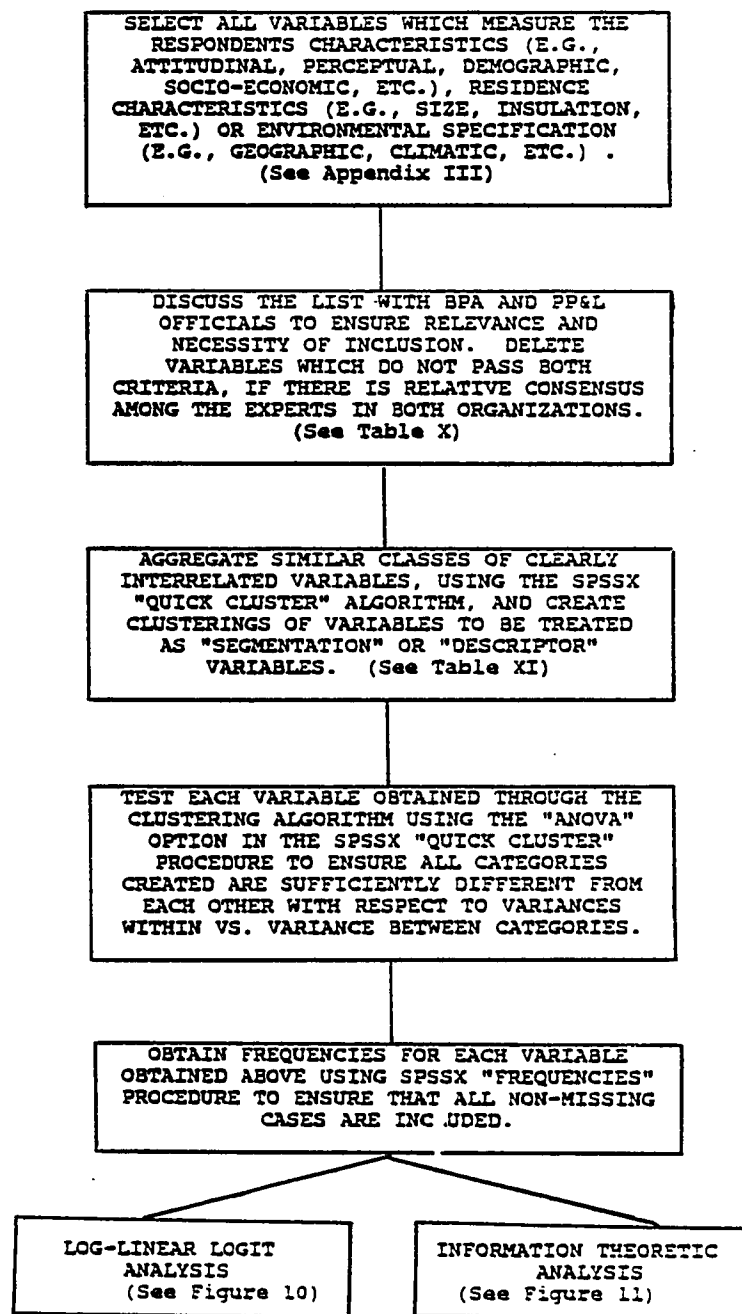


Figure 9. Flow chart for selection and aggregation of research variables

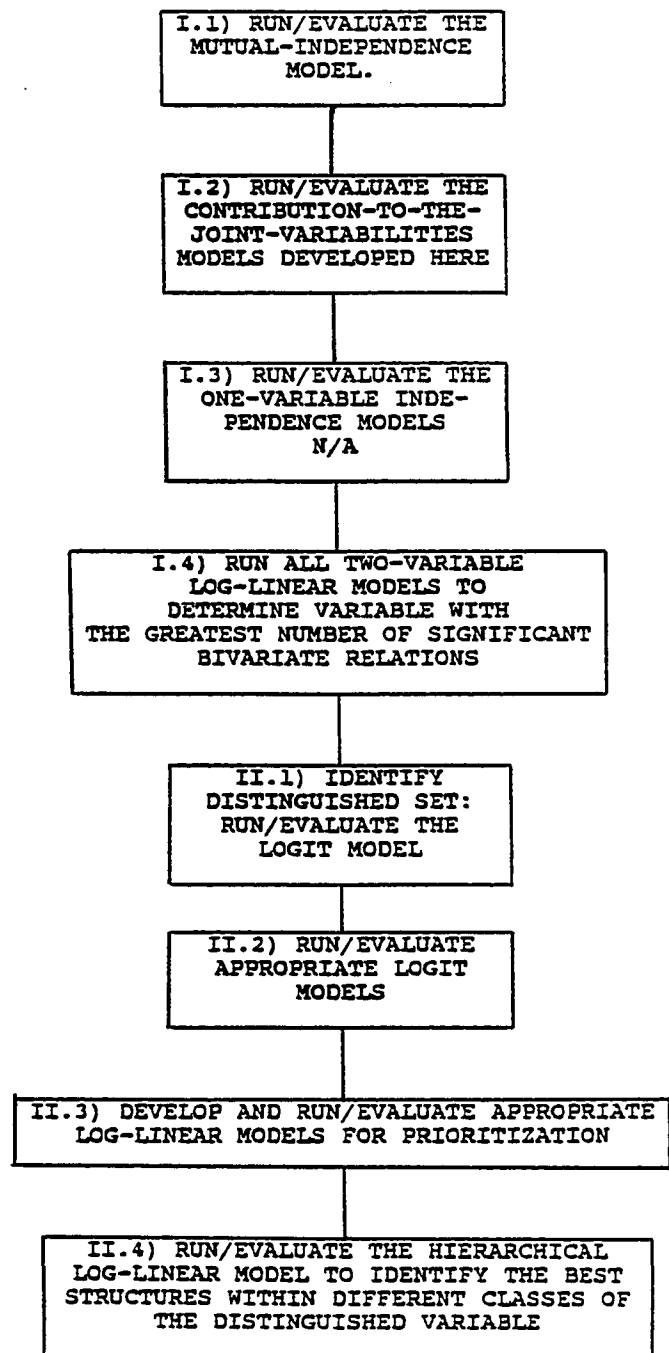


Figure 10. Flow chart of the log-linear/logit analysis.

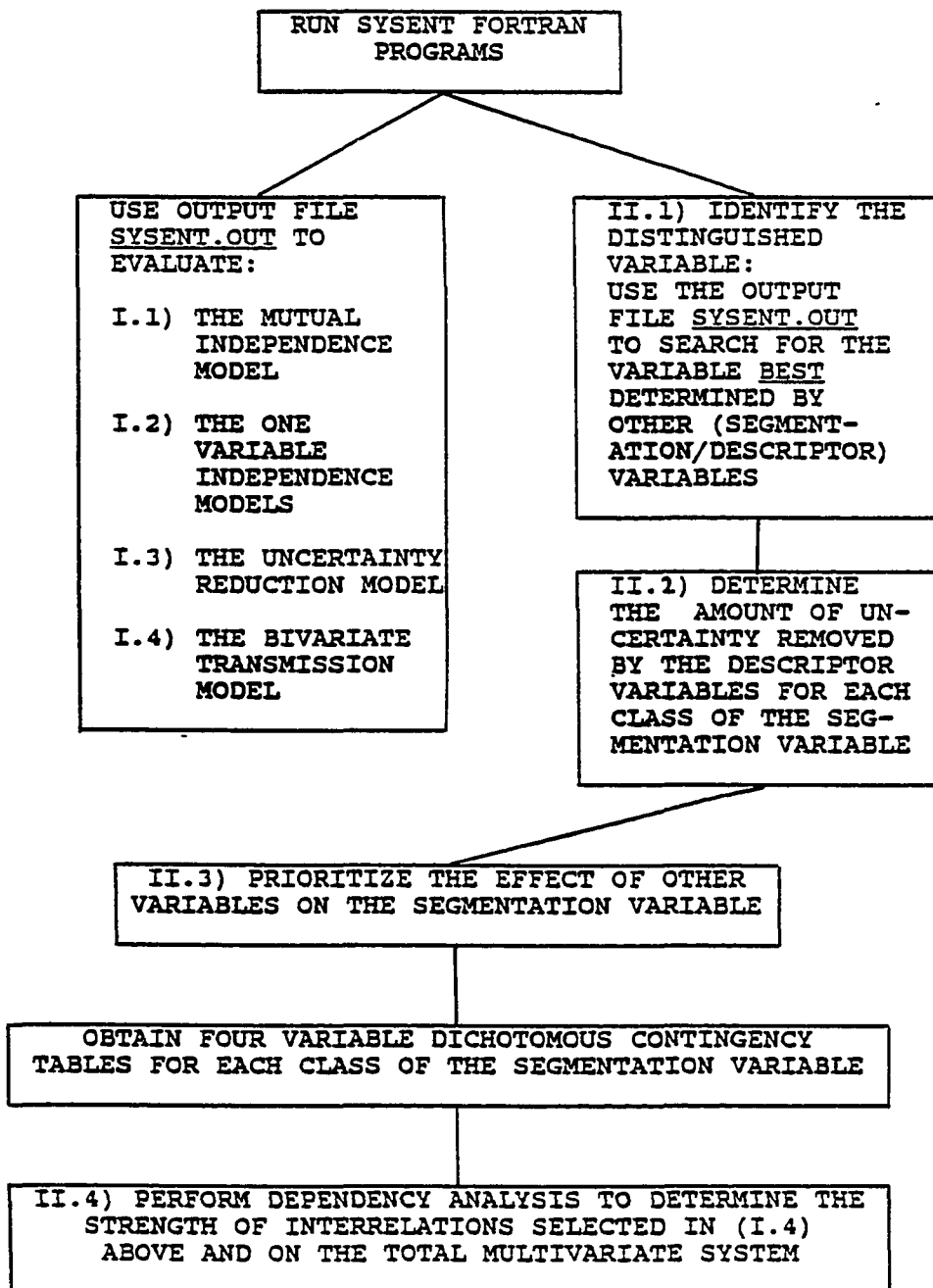


Figure 11. Flow chart of the information theoretic analysis

TABLE X
LIST OF VARIABLES PRIOR TO AGGREGATIONS

VARIABLE NAME	VARIABLE LABEL	QUESTION NUMBER
<u>GENATT</u>		
VAR049	Energy Efficiency of home	20
VAR050	Thought on environmental pollution	20
VAR051	Thoughts on cost of energy	20
VAR052	Thoughts on unemployment	20
VAR053	Thoughts on scarcity of energy	20
VAR054	Thoughts on inflation	20
VAR055	Thoughts on Crime	20
<u>ENRATT</u>		
VAR067	Statement-reducing water temp. saves	22
VAR069	Statement-turning down temp. worthwhile	22
VAR073	statement-conserving energy=change in life	22
<u>BEHAVE</u>		
VAR075	Made use of federal tax credit	24
VAR077	Made use of state tax credit	24
VAR079	Made use of low interest loan	24
<u>PERCEPT</u>		
VAR066	Statement-comfortable at or < 68F	22
VAR071	Statement-right to use energy	22
VAR072	Statement-price more important than energy	22
<u>AWARE</u>		
VAR074	Heard of Federal tax credit	22
VAR076	Heard of State tax benefit	22
VAR078	Heard of low interest loan	22
<u>KWHUSE</u>		
KWH	Total kilo-watt-hours used in 12 months	*
<u>CLIMGEO</u>		
GEOAREA	Geographic area	**
<u>TYPDWEL</u>		
VAR005	Type of dwelling	2
VAR469	Total square feet first floor	**
VAR470	Total square feet second floor	**
VAR471	Total square feet third floor	**
<u>RENTOWN</u>		
VAR008	Means of payment for housing	5
<u>DEMOG</u>		
VAR318	Total number of residents	129
VAR360	Ethnic origin	132
VAR362	Level of education	133
VAR364	Combined 1982 income	134
<u>INSUL</u>		
VAR133	Percent insulation, roof	34
VAR141	Percent insulation, floor	41

* Merged from a different questionnaire

** No question number listed in the questionnaire.

The SPSSX QUICK CLUSTER procedure was utilized to aggregate each of the above groups of variables. These clusters were named GENATT, ..., ENRATT, attitudinal cluster, PERCEPT, perceptual cluster, AWARE, awareness cluster, and BEHAVE, electricity usage behavior.

The study began by identifying a set of segmentation variables in two separate stages. In the first stage the a priori method was used to identify a large number of variables. Next, the clustering approach was used to group these variables into reasonable clusters identify the segmentation and descriptor variables. Then again, Phase I and II of research questions were addressed. The results of this study establish a set of market segments useful to energy planners and other decision makers in the field of utilities and power administration.

Chapter IV presents analyses and results of the PNWRES study through the two main approaches (i.e., information theoretic and log-linear methods).

TABLE XI
LIST OF VARIABLES (OR CLUSTERS) AFTER
AGGREGATION

CLUSTER/ VARIABLE	DESCRIPTION	VARIABLES CLUSTERED (SEE APPENDIX IV)
<u>A. SEGMENTATION CLUSTERS:</u>		
GENATT	General attitudes (economy and environment)	VAR049 THRU VAR055
ENRATT	Energy related attitudes	VAR067, VAR069, VAR073
PERCEPT	Energy (comfort/discomfort) perceptions	VAR066, VAR071, VAR072
AWARE	Energy-saving-development funding awareness	VAR074, VAR076, VAR078
BEHAVE	Energy consumption behavior	VAR075, VAR077, VAR079
KWHUSE	Energy usage in kilo-watt- hours	KWH
<u>B. DESCRIPTOR CLUSTERS:</u>		
CLIMGEO	Climatic/geographic environment	GEOAREA
TYPDWEL	Type of dwelling	VAR005 VAR469, VAR470 VAR471
RENTOWN	Status of the resident in terms of renter/owner	VAR008
DEMOG	Demographic cluster	VAR318, VAR360 VAR362, VAR364
INSUL	Level of insulation of the residence	VAR133, VAR141

CHAPTER IV

RESULTS

This chapter presents the results of the segment congruence analysis of PNWRES data. First, the research questions for Phase I are addressed. In this phase the segmentation variables are analyzed to:

1. Determine the existence and amount of the total mutual association among segmentation variables.
2. Identify the variable(s) with the highest contribution toward the total mutual association.
3. Identify the variable(s) with the highest contribution toward the joint variability in the system of segmentation variables.
4. Identify the variable(s) which result in the highest amount of reduction in the uncertainty of the system of segmentation variables.
5. Identify the variable(s) with the greatest number of significant dyadic relations with the other variables.

Phase II, the selection and analysis of the distinguished segmentation variable, is addressed next. The other segmentation variables are used as descriptor

variables in one set of analyses, while five exogenous descriptor variables are used in another, in order to:

1. Identify the distinguished (i.e., the best) segmentation variable.
2. Assess the changes in the reduction of uncertainty in the distinguished variable, due to the descriptor variables, under different collapsing schemes.
3. Prioritize the contributions of the descriptor variables to the determination of the distinguished variable.
4. Analyze the structure of the distinguished variable and all of the descriptor variables using dependency analysis.

The energy usage variable (KWHUSE) is of particular importance to the BPA decision makers, as well as the local utilities. Therefore, this variable will be included as a distinguished variable in addition to the analytically selected distinguished variable, should they not be the same.

In order to provide a comparative base with the classical segment congruence analysis, Phase II is first implemented using the endogenous variables (i.e., other segmentation variables) as descriptor variables, with the a priori selected, electricity consumption as the distinguished variable. Next, the analytically derived distinguished variable is analyzed. (It will be seen that energy related behavior is selected analytically.)

Phase II is then repeated using exogenous descriptor variables to identify the distinguished variable and proceed with its analysis. (Here electricity consumption is selected analytically.) Figure 12 shows the flow of these analyses:

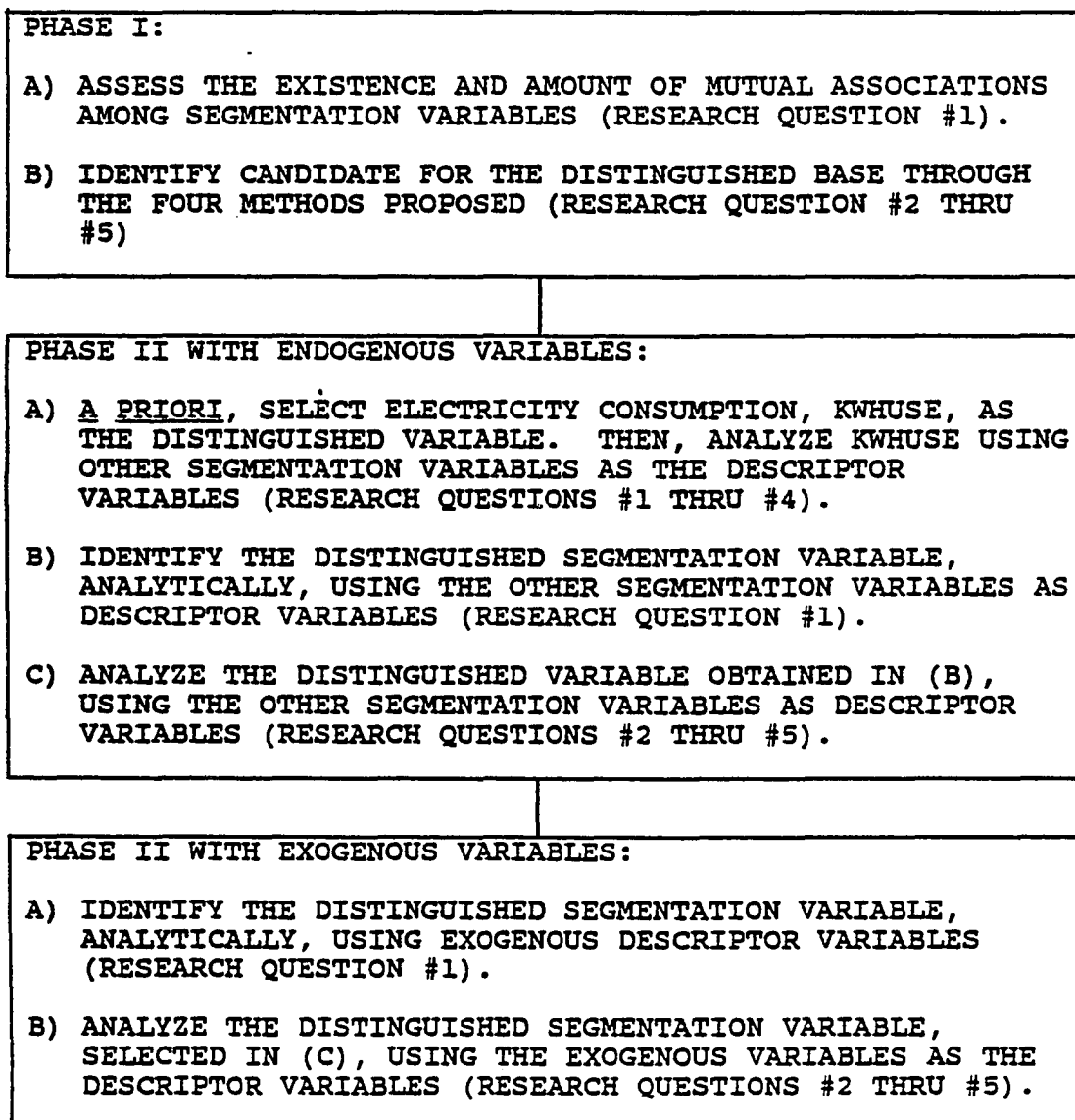


Figure 12. Condensed flow chart of data analysis.

ANALYSIS OF THE SEGMENTATION VARIABLES

In addition to assessing the mutual association among the segmentation variables, this section identifies a set of candidate segmentation variables using various methods. In cases where a variable is selected as the candidate base by two different methods, as a convention, the second best variable will be designated as the candidate. This will ensure that each method will produce a new candidate. The logic for this convention relates to the fact that segment congruence analysis is ideally performed on a large number of variables, therefore, retaining at least four variables for further analyses is reasonable.

Existence and Degree of Mutual Association Among Segmentation Variables

First a set of segmentation variables (or clusterings) are introduced. Table XI in chapter III presents these variables, and Table X lists the variables selected from the PNWRES data (see Appendix IV for a complete list of variables). Chapter III also describes procedures used to obtain and cluster all of the variables used in this study. Table XII reproduces the listing of the segmentation variables for ease of reference.

TABLE XII
LIST OF THE SEGMENTATION VARIABLES

Variable (Abbreviation)	Description	# of Classes
GENATT (G)	General Attitudes	4
ENRATT (E)	Energy Related Attitudes	5
PERCEPT (P)	Energy Related Perceptions	4
AWARE (A)	Energy Related Awareness	3
BEHAVE (B)	Energy Related Behavior	2
KWHUSE (K)	Electricity Consumption	5

The analyses in Phase I will use the above segmentation variables. The first research question involves the assessment of mutual association among these variables.

Research Question #1. The hypothesis tested for the mutual association assessment is as follows:

H_0 : All of the variables in the system of segmentation variables are independent.

H_A : Two or more of the segmentation variables are interdependent

Table XIII presents the results of 1) the independence model based on the general log-linear approach, and 2) the transmission-based analysis of the mutual association among the segmentation variables.

TABLE XIII

ASSESSMENT OF THE MUTUAL ASSOCIATION

Variables: GENATT (4), ENRATT (5), PERCEPT (4), AWARE (3),
 BEHAVE (2), KWHUSE (5),
 where; (n) = Corresponding variable has n Classes.

LOG-LINEAR APPROACH

Model: {GENATT}{ENRATT}{PERCEPT}{AWARE}{BEHAVE}{KWHUSE}

L^2_{LL}	df	p	N
3150.14	2382	≈ 0	3335

INFORMATION THEORETIC APPROACH

Model: T(GENATT:ENRATT:PERCEPT:AWARE:BEHAVE:KWHUSE)

T	T/T _{MAX}	L^2_{LL}	df	p	N
0.6814	.0765	3150.14	2382	≈ 0	3335

Examining Table XIII reveals that both the general log-linear model and the information theoretic model reject the null hypothesis with likelihood-ratio chi-square of 3150.14, df= 2382, and $p \approx 0$. The transmission measure, T, of .6814, and T/T_{max} of .0765 indicate that the amount of

overlap is relatively small (but highly significant, statistically).

Thus, it is concluded that general attitudes, energy related attitudes, perceptions, awareness, and behavior, and energy consumption are mutually associated. This indicates that knowledge about the class of one or more of the variables reveals information about the classes of the other variables.

The remaining research questions in this phase, aim at determining which variables are likely candidates for a distinguished segmentation variable. Four different methods are proposed: Contribution of each segmentation variable to the (a) total mutual association, (b) the joint variability in the system, and (c) total uncertainty in the system; and (4) the number of significant (dyadic) relations between variables associated with each candidate variable.

Identification of Candidates for the Distinguished Variables

Research Question #2. Following is the list of the hypotheses tested in order to assess contribution of individual variables to the total mutual association among the segmentation variables:

- 1) H_0 : GENATT does not contribute toward the total mutual association in the system of segmentation variables.
- H_A : GENATT contributes toward the total mutual association in the system of segmentation variables.

- 2) H_0 : ENRATT does not contribute toward the total mutual association in the system of segmentation variables.
 H_A : ENRATT contributes toward the total mutual association in the system of segmentation variables.
- 3) H_0 : PERCEPT does not contribute toward the total mutual association in the system of segmentation variables.
 H_A : PERCEPT contributes toward the total mutual association in the system of segmentation variables.
- 4) H_0 : AWARE does not contribute toward the total mutual association in the system of segmentation variables.
 H_A : AWARE contributes toward the total mutual association in the system of segmentation variables.
- 5) H_0 : BEHAVE does not contribute toward the total mutual association in the system of segmentation variables.
 H_A : BEHAVE contributes toward the total mutual association in the system of segmentation variables.
- 6) H_0 : KWHUSE does not contribute toward the total mutual association in the system of segmentation variables.
 H_A : KWHUSE contributes toward the total mutual association in the system of segmentation variables.

Table XIV shows the results of the log-linear one-variable-independence models and their information theoretic equivalent models:

TABLE XIV
ASSESSMENT OF THE CONTRIBUTION OF EACH VARIABLE TO THE
MUTUAL ASSOCIATION

Variables: GENATT (4), ENRATT (5), PERCEPT (4), AWARE (3),
BEHAVE (2), KWHUSE (5),
where: (n) = Corresponding variable has n Classes.

LOG-LINEAR APPROACH

Models: E.G.,

{GENATT, ENRATT} {GENATT, PERCEPT} {GENATT, AWARE} {GENATT, BEHAVE}
{GENATT, KWHUSE}

(1) Base (i)	(2) L^2_{LL}	(3) df	(4) p
GENATT	2884.37	2340	= 0
ENRATT	3013.86	2330	= 0
PERCEPT	2861.70	2340	= 0
AWARE	1269.70	2352	1
BEHAVE	1279.92	2366	1
KWHUSE	3068.89	2330	= 0

INFORMATION THEORETIC APPROACH

Models: E.G.,

$\Phi_1 T = T_{\text{GENATT}}(\text{ENRATT:PERCEPT:AWARE:BEHAVE:KWHUSE})$

(1) Base (i)	(2) T_i	(3) $1-T_i/T$	(4) L^2_{IT}	(5) df	(6) p
GENATT	0.6239	0.0844	2884.29	2340	= 0
ENRATT	0.6519	0.0433	3013.74	2330	= 0
PERCEPT	0.6190	0.0916	2861.64	2340	= 0
AWARE	0.2747	0.5969	1269.84	2352	1
BEHAVE	0.2768	0.5938	1279.55	2366	1
KWHUSE	0.6637	0.0260	3068.29	2330	= 0

The top part of Table XIV (i.e., the log-linear approach), includes the segmentation variables under the consideration, column 1, and their corresponding likelihood ratio chi-square, degrees of freedom, and risk levels in columns 2, 3, and 4, respectively. The bottom part of this table, on the other hand, presents the information theoretic models equivalent to the log-linear models.

It is seen that both models generate identical statistics, however, the information theoretic model generates a measure of the amount contribution (i.e., T_i in column 2), which is not given in the log-linear approach. Also, column 3 gives a normalized measure of the amount of contribution toward the joint variability in the system ($1 - T_i/T$) which has a maximum of one (i.e., when all of the association is due the variable under the consideration), and a minimum of zero (i.e., when the variable under investigation is independent of others).

Columns 4, 5, and 6 present the likelihood-ratio chi-square, degrees of freedom, and the risk levels for each model. L^2_{IT} is calculated as $2N\log(2)T$, and degrees of freedom are calculated based on the formula given by Broekstra (1981) and reproduced in the methodology chapter. These calculations give the same values as those produced by the log-linear model. The differences in the decimals are attributed to the fact that cells with sampling zeros had to

be given a negligible value (i.e., $1E-20$) in the log-linear model, as the proportional fitting method used would generate incorrect statistics without such adjustment.

In order to select the best variable the magnitude of the likelihood-ratio chi-square was used for the log-linear approach, and the normalized measure in column 3 (i.e., $1 - T_i/T$) was used for the information theoretic approach. Based on both approaches, the energy related awareness variable, AWARE, is selected as the candidate for the distinguished segmentation variable.

Research Question #3. This research question assesses the contribution of each of the segmentation variables toward the joint variability in the system of segmentation variables. In information theoretic terms, this assessment leads to the identification of the variable(s) with the highest contribution toward the total systematic entropy in the system.

Following is the list of the hypotheses tested in order to assess the second research question in this phase:

- 1) H_0 : GENATT does not contribute toward the joint variability in the system of segmentation variables.
 H_A : GENATT contributes toward the joint variability in the system of segmentation variables.
- 2) H_0 : ENRATT does not contribute toward the joint variability in the system of segmentation variables.

- H_A : ENRATT contributes toward the joint variability in the system of segmentation variables.
- 3) H_O : PERCEPT does not contribute toward the joint variability in the system of segmentation variables.
- H_A : PERCEPT contributes toward the joint variability in the system of segmentation variables.
- 4) H_O : AWARE does not contribute toward the joint variability in the system of segmentation variables.
- H_A : AWARE contributes toward the joint variability in the system of segmentation variables.
- 5) H_O : BEHAVE does not contribute toward the joint variability in the system of segmentation variables.
- H_A : BEHAVE contributes toward the joint variability in the system of segmentation variables.
- 6) H_O : KWHUSE does not contribute toward the joint variability in the system of segmentation variables.
- H_A : KWHUSE contributes toward the joint variability in the system of segmentation variables.

As mentioned previously, the single-variable independence model in the log-linear approach has been misinterpreted as assessing the contribution of individual variables toward the total mutual association. It has been shown in the present study that this model assesses instead the contribution of individual variables to the joint variability in the multivariate system.

Table XV shows the results, based on this new interpretation of the log-linear one-variable-independence models, and their information theoretic equivalent models:

The log-linear portion in Table XV includes the segmentation variables under the consideration, column 1, and their corresponding likelihood ratio chi-square, degrees of freedom, and risk levels in columns 2, 3, and 4, respectively. The bottom part of this table, on the other hand, presents the information theoretic equivalent of this model.

Again, both models generate identical statistics, however, the information theoretic model generates a measure of the amount contribution (i.e., $S-S_i$ in column 2), which is not given in the log-linear approach. Also, column 3 gives the normalized measure of the amount of contribution toward the joint variability in the system.

In order to determine which variable contributes the most toward the joint variability in the system, Green and Carmone's (1977) logic is followed. A model which gives a lower L^2 fits the data better than a model that gives a higher L^2 . Therefore, for the single-variable-independence model, the model that gives the highest L^2 should be selected as the model which fits the data the least.

TABLE XV
ASSESSMENT OF THE CONTRIBUTION OF EACH
VARIABLE TO THE JOINT VARIABILITY IN THE
SYSTEM OF SEGMENTATION VARIABLES

Variables: GENATT (4), ENRATT (5), PERCEPT (4), AWARE (3),
BEHAVE (2), KWHUSE (5),
where: (n) = Corresponding variable has n Classes.

LOG-LINEAR APPROACH

Models: E.G.,

$$\begin{aligned} &(\text{GENATT})(\text{ENRATT}, \text{PERCEPT}, \text{AWARE}, \text{BEHAVE}, \text{KWHUSE})) = \\ &((\text{GENATT})(\text{ENRATT})(\text{PERCEPT})(\text{AWARE})(\text{BEHAVE})(\text{KWHUSE}))) - \\ &((\text{ENRATT})(\text{PERCEPT})(\text{AWARE})(\text{BEHAVE})(\text{KWHUSE})) \end{aligned}$$

(1) Base (i)	(2) L^2_{LL}	(3) df	(4) p
GENATT	858.53	1797	= 0
ENRATT	809.31	1916	= 0
PERCEPT	861.31	1797	= 0
AWARE	2320.77	1598	= 1
BEHAVE	2111.75	1199	= 1
KWHUSE	724.45	1916	= 0

INFORMATION THEORETIC APPROACH

Models: E.G.,

$$\begin{aligned} \Phi_1 S &= S(\text{GENATT:ENRATT:PERCEPT:AWARE:BEHAVE:KWHUSE}) - \\ &S_{\text{GENATT}}(\text{ENRATT:PERCEPT:AWARE:BEHAVE:KWHUSE}) = \\ &T(\text{GENATT:ENRATT:PERCEPT:AWARE:BEHAVE:KWHUSE}) - \\ &T(\text{ENRATT:PERCEPT:AWARE:BEHAVE:KWHUSE}) \end{aligned}$$

(1) Base (i)	(2) $S-S_i$	(3) $1-S_i/S$	(4) L^2_{IT}	(5) df	p
GENATT	0.1857	0.1893	858.53	1797	= 0
ENRATT	0.1751	0.1785	809.31	1916	= 0
PERCEPT	0.1863	0.1899	861.31	1797	= 0
AWARE	0.5020	0.5117	2320.77	1598	= 1
BEHAVE	0.4568	0.4656	2111.75	1199	= 1
KWHUSE	0.1567	0.1597	724.45	1916	= 0

Based on the above reasoning, both the log-linear and the information theoretic approach would result in the choice of the energy related awareness (AWARE) as a candidate for the distinguished segmentation variable. Based on the convention established at the beginning of this section, however, the second best variable (i.e., energy use behavior, BEHAVE) will be selected (because, AWARE has already been selected). Therefore, the variable BEHAVE is designated as the candidate segmentation variable.

Research Question #4. This research question examines the contribution of each variable toward the overall uncertainty in the system. First, the significance of each contribution is tested, and next, the amount of contribution is used to identify a candidate for the distinguished variable.

Following is the list of hypotheses tested to address the significance of each variable's contribution:

- 1) H_0 : GENATT does not contribute toward the overall uncertainty in the system of segmentation variables.
 H_A : GENATT contributes toward the overall uncertainty in the system of segmentation variables.
- 2) H_0 : ENRATT does not contribute toward the overall uncertainty in the system of segmentation variables.
 H_A : ENRATT contributes toward the overall uncertainty in the system of segmentation variables.

- 3) H_O : PERCEPT does not contribute toward the overall uncertainty in the system of segmentation variables.
- H_A : PERCEPT contributes toward the overall uncertainty in the system of segmentation variables.
- 4) H_O : AWARE does not contribute toward the overall uncertainty in the system of segmentation variables.
- H_A : AWARE contributes toward the overall uncertainty in the system of segmentation variables.
- 5) H_O : BEHAVE does not contribute toward the overall uncertainty in the system of segmentation variables.
- H_A : BEHAVE contributes toward the overall uncertainty in the system of segmentation variables.
- 6) H_O : KWHUSE does not contribute toward the overall uncertainty in the system of segmentation variables.
- H_A : KWHUSE contributes toward the overall uncertainty in the system of segmentation variables.

In Chapter 3, it was argued that the application of the log-linear approach to research questions which deal with the uncertainty in a multivariate system is not necessary (i.e., due to eventual necessity to resort to the entropy measure and the lack of availability of a known computer program to perform such analysis). Therefore, the log-linear approach will not be applied for this question as well as other questions dealing with overall system uncertainty. Table XVI presents the results of the information theoretic analysis:

TABLE XVI
ASSESSMENT OF THE CONTRIBUTION OF EACH VARIABLE
TO THE TOTAL UNCERTAINTY IN THE SYSTEM OF
SEGMENTATION VARIABLES

Models: E.G.,

$$\Phi_1 H = H(\text{GENATT}, \text{ENRATT}, \text{PERCEPT}, \text{AWARE}, \text{BEHAVE}, \text{KWHUSE}) -$$

$$H(\text{GENATT})$$

INFORMATION THEORETIC APPROACH

(1)	(2)	(3)	(4)	(5)	(6)	(7)
i	H(i)	H _i	% Removed	L ²	d.f.	p
GENATT	1.5571	5.9859	20.64	2047.66	3	0
ENRATT	2.2413	5.3017	29.71	372.77	4	0
PERCEPT	1.3687	6.1743	18.15	2918.68	3	0
AWARE	1.0131	6.5300	13.43	2643.89	2	0
BEHAVE	0.6186	6.9245	8.20	1763.32	1	0
KWHUSE	1.4256	6.1175	18.90	4143.99	4	0

In Table XVI, Column 1, lists the variables whose contribution to the overall uncertainty in the system is being investigate. Column 2, H(i), is the uncertainty of each variable i. Column 3, H_i, is the uncertainty in the overall system given the variable i (i.e., the remaining uncertainty). Column 4, % removed, is H(i)/(total uncertainty in the system). Column 5, is the L² calculated

based on the relationship between uncertainty and likelihood-ratio chi-square (i.e., $L^2 = 2N \log(2)H(i) - 2N \log A$, where N =total number of observations, $H(i)$ =entropy of variable i , and A =number of states of variable i). For instance, the L^2 for i =GENATT is calculated as:

$$L^2 = 2*3335*\log(2)*1.5571 - 2*3335*\log(4) = 2047.66.$$

Columns 6 and 7 list the degrees of freedom and the risk level for each variable under consideration.

Based on the L^2 , all of these variables contribute significantly to the overall uncertainty in the system ($p=0$ for rejecting the null hypotheses for all variables). Beyond this result no other interpretations can be made as the degrees of freedom differ in each case and therefore the L^2 are not directly comparable.

Judging by the percent of contribution of each variable, the energy related attitudes variable, ENRATT is selected as the candidate for the distinguished segmentation variable (ENRATT contributes 29.71% to the overall uncertainty).

Research Question #5. In order to address this research question, bivariate associations were investigated for all possible two variable combinations, using both methods. The likelihood-ratio chi-square was then used to assess the significance of each bivariate model. For this purpose, a two dimensional table was constructed of the

likelihood-ratio chi-squares and, for the information theoretic approach, the amounts of bivariate transmission.

To assess the significance of the association within each two-variable combination, hypotheses similar to the following were tested:

H_0 : The two variables being considered are independent.

H_A : The two variables are interdependent

Table XVII presents the two bivariate tabulations. One based on the likelihood-ratio chi-squares derived from the log-linear models, and another based the information theoretic bivariate transmissions and likelihood-ratio chi-squares derived from these transmissions:

The bivariate table for the log-linear approach includes the likelihood-ratio chi-square, the degrees of freedom, and the risk value for each bivariate combination of segmentation variables. These figures appear in the upper triangle of the top bivariate table. The information theoretic approach produces the same bivariate table, with identical upper triangle, but with the addition of measures of association (i.e., bivariate transmissions), presented in the lower triangle.

TABLE XVII
ASSESSMENT OF THE NUMBER OF SIGNIFICANT
DYADIC RELATIONS

LOG-LINEAR APPROACH

Models: E.G.,

(GENATT,ENRATT), (GENATT,PERCEPT), (GENATT,AWARE), etc.

	GENATT	ENRATT	PERCEPT	AWARE	BEHAVE	KWHUSE
GENATT						
L^2_{LL}	-	36.05*	200.64*	11.04	11.77	6.27
df	-	12	9	6	3	12
P	-	.000	.000	.087	.008	.902
ENRATT						
L^2_{LL}		-	24.45	24.06*	16.41*	35.31*
df		-	12	8	4	16
P		-	.018	.002	.003	.004
PERCEPT						
L^2_{LL}			-	22.92*	13.09*	27.34
df			-	6	3	12
P			-	.001	.004	.007
AWARE						
L^2_{LL}				-	1819.29*	2.92
df				-	2	8
P				-	.000	.939
BEHAVE						
L^2_{LL}					-	9.91
df					-	4
P					-	.042
KWHUSE						
L^2_{LL}						-
df						-
P						-

INFORMATION THEORETIC APPROACH

Models: E.G.,

T(GENATT:ENRATT), T(GENATT:PERCEPT), T(GENATT:AWARE), etc.

	GENATT	ENRATT	PERCEPT	AWARE	BEHAVE	KWHUSE
GENATT						
L^2_{LL}	-	36.05*	200.64*	11.04	11.77	6.27
df	-	12	9	6	3	12
P	-	.000	.000	.087	.008	.902
ENRATT						
L^2_{LL}	0.0078	-	24.45	24.06*	16.41*	35.31*
df		-	12	8	4	16
P		-	.018	.002	.003	.004
PERCEPT						
L^2_{LL}	0.0434	0.0053	-	22.92*	13.09*	27.34
df			-	6	3	12
P			-	.001	.004	.007
AWARE						
L^2_{LL}	0.0024	0.0052	0.005	-	1819.29*	2.92
df				-	2	8
P				-	.000	.939
BEHAVE						
L^2_{LL}	0.0025	0.0035	0.0028	0.3935	-	9.91
df					-	4
P					-	.042
KWHUSE						
L^2_{LL}	0.0014	0.0076	0.0059	0.0006	0.0021	-
df						-
P						-

Judging on the basis of significance of the bivariate likelihood-ratio chi-square, with the confidence level at 0.005, the significant bivariate associations are designated with "*" in both tables. It is seen that the energy related attitudes variable, ENRATT, has the highest number of bivariate associations. However, this variable has already been selected, therefore, the energy related perception variable, PERCEPT, is selected as the candidate variable.

Note that, there are two other variables (i.e., AWARE and BEHAVE) with the same number of significant bivariate associations, which have already been selected. In case of a tie between two (or more) variables, both (all) of which have not been selected yet, the confidence level is set at progressively narrower levels until one of the variables prevails. In this application, however, this was not necessary and PERCEPT was selected as the candidate for the distinguished variable.

DETERMINATION AND ANALYSIS OF THE DISTINGUISHED SEGMENTATION VARIABLE USING THE OTHER SEGMENTATION VARIABLES AS THE DESCRIPTOR VARIABLES

In Phase I, four candidates for the distinguished segmentation variable were identified. These variables include ENRATT, PERCEPT, AWARE, BEHAVE AND KWHUSE. Phase II continues the process of segment congruence analysis by assessing predictability of various segmentation variables

by a set of descriptor variables (i.e., either endogenous, using other segmentation variables, or exogenous, using other descriptor variables, such as demographics, socioeconomics, etc.).

First, electricity consumption, KWHUSE, is selected as the distinguished variable, a priori. This variable is then analyzed using the other segmentation variables (i.e., endogenous variables) as descriptor variables. This will virtually replicate Green and Carmone (1977) analysis with the addition of other research questions introduced in the present study.

The analyses will then proceed by identifying and analyzing the distinguished segmentation variable using the endogenous variables first. Then, in the section 5.3, a set of exogenous variables will be used in order to identify and analyze the distinguished variable.

Electricity Consumption as the Distinguished Segmentation Variable (A Priori)

Research Question #1. Electricity consumption, KWHUSE, was identified by the BPA and PP&L executives as the most significant variable. In the context of market segmentation, also, this variable is of particular interest, as it is the variable by which the heavy user group can be distinguished from the light user. Therefore, KWHUSE is selected, a priori, as the distinguished segmentation variable in part (A) of this section.

Research Question #2. This research question aims at measuring the amount of reduction in uncertainty in the distinguished variable when different classes are re-clustered (e.g., into dichotomies of a cluster of interest and all other clusters). The segmentation variable (here, energy consumption, KWHUSE) is examined with respect to three different groupings: the low-consumption group (classes 1 & 2 of KWHUSE), the medium-consumption group (class 3 of KWHUSE), and the high-consumption group (classes 4 & 5 of KWHUSE). Table XVIII shows the results of these analyses.

In Table XVIII, the logit analysis for each reclassification scheme, column 1, produces the goodness-of-fit likelihood ratio chi-square, column 2, and its associated degrees of freedom, and risk level, columns 3 and 4, respectively. Additionally, the logit model provides the measures of entropy and concentration, which are measures of association between the dependent variable and independent variables (refer to the chapter II, literature search, pp. 35-36). These measure are shown in columns 5 and 6, respectively.

Column 7 shows the ranking of different reclassifications based on their measure of entropy.

TABLE XVIII
PREDICTABILITY OF A PARTICULAR CLASS VS. ALL
OTHER CLASSES OF THE DISTINGUISHED VARIABLE

<u>LOG-LINEAR APPROACH (THE LOGIT MODEL)</u>						
(1) Reclassi- fication Scheme for KWHUSE	(2) L^2_{LL}	(3) df	(4) p	(5) <u>Ana. of Disper.</u> Entropy	(6) Concent	(7) Rank
Class 1&2 = 1 Others = 2	218.45	466	1	0.0180	0.0122	2
Class 3 = 1 Others = 2	250.02	466	1	0.0051	0.0069	3
Class 4&5 = 1 Others = 2	212.16	466	1	0.0111	0.0134	1
<u>INFORMATION THEORETIC APPROACH</u>						
(1) Reclassi- fication Scheme for KWHUSE	(2) Uncert. of seg. Var. H(K)	(3) Uncert of seg.Var. given others H _. (K)	(4) Amount (%) Uncert. Predicted	(5) Rank		
Class 1&2 = 1 Others = 2	0.8596	0.8042	0.0554 (6.45%)	2		
Class 3 = 1 Others = 2	0.9675	0.9085	0.0590 (6.10%)	3		
Class 4&5 = 1 Others = 2	0.5037	0.4474	0.0563 (11.18%)	1		

Ranking based on only likelihood-ratio chi-square would also produce the same ranking, however, it is customary to rank on the basis of a measure of association rather than a measure of statistical significance. Furthermore, the measure of entropy was selected because it assesses the amount of total variability due to (or accounted for by) the independent variable. Based on this ranking, the high electricity consumption group is predicted better than the other reclassifications ($L^2=212.16$) followed by the low and the medium-consumption groups ($L^2=218.45$ and 250.02 , respectively). As can be seen here, in logit analysis, lower amounts of L^2 indicated a better fit for the logit function than the higher ones (see Green and Carmone 1977).

The information theory model, in Table XVIII, measures (in column 2) the uncertainty in each reclassification (as represented by the entropy of the variable), the uncertainty of that variable, given the descriptor variables, column 3, the amount and the percent of uncertainty removed by the descriptor variables, column 4.

Column 5 gives the ranking of the predictive power based on the information theoretic model. The ranking is identical to that produced by the logit model. This is not a coincidence, since, the information theoretic model addresses the amount of variability remaining in the criterion variable data, with the knowledge of a set of independent variables. (This is reminiscent to the measure

of entropy.) The entropy of the dependent variable knowing the independent variables, in essence measures the residuals in the logit model.

Therefore, it is concluded that the low-consumption group is best predicted by the descriptor variables. It is also seen that, the medium-consumption group is determined by the least amount by the other variables.

Research Question #3. In order to prioritize the contributions of independent variables toward the determination of the energy usage variable, KWHUSE, the general log-linear model was re-configured and applied. The upper part of Table XIX, presents the results.

Abrahams and van Bueren's (1980) procedure was used for the ranking of the descriptor variables using information theory. The bottom part of Table XIX shows this ranking. The first step evaluates the bivariate marginal transmissions (here, ENRATT is selected as the variable with the highest bivariate transmission). Next, bivariate conditional (partial) transmission (on the condition that ENRATT is given/known), and so on.

This procedure produced the following ranking: ENRATT, PERCEPT, GENATT, AWARE, and BEHAVE. The two approaches produced the same ranking. This same procedure is used by Conant (1982) in developing the dependency analysis. This technique is applied in Research Question #4.

TABLE XIX

PRIORITIZED EFFECT OF THE OTHER SEGMENTATION
VARIABLES ON ENERGY CONSUMPTION

LOG-LINEAR APPROACH

Log-Linear Model	L^2_{LL}	df	p
(K)(G)	6.27	12	0.902
(K)(E)	35.31*	16	0.004
(K)(P)	27.34	12	0.007
(K)(A)	2.92	8	0.939
(K)(B)	9.91	4	0.042
(K)(EG)	102.28	76	0.024
(K)(EP)	110.86*	76	0.006
(K)(EA)	73.68	56	0.057
(K)(EB)	59.29	36	0.009
(K)(EPG)	278.95*	316	0.934
(K)(EPA)	228.93	236	0.617
(K)(EPB)	177.25	156	0.117
(K)(EPGA)	563.36*	956	1.000
(K)(EPGB)	437.19	642	1.000
(K)(EPGAB)	724.45*1916		1.000

INFORMATION THEORETIC APPROACH

Model	T	L^2_{IT}	df	p	Cum L^2_{IT}
T(K:G)	0.0014	6.47	12	0.902	
T(K:E)	0.0076*	35.14	16	0.004	35.14
T(K:P)	0.0059	27.28	12	0.004	
T(K:A)	0.0006	2.77	8	0.939	
T(K:B)	0.0021	9.71	4	0.042	
$T_E(K:G)$	0.0145	67.04	60	0.024	
$T_E(K:P)$	0.0163*	75.36	60	0.006	110.50
$T_E(K:A)$	0.0083	38.37	40	0.057	
$T_E(K:B)$	0.0052	24.04	20	0.009	
$T_{EP}(K:G)$	0.0364*	168.29	240	0.934	278.79
$T_{EP}(K:A)$	0.0255	117.89	160	0.617	
$T_{EP}(K:B)$	0.0144	66.58	80	0.117	
$T_{EPG}(K:A)$	0.0616*	284.80	640	1.000	563.59
$T_{EPG}(K:B)$	0.0342	158.12	320	1.000	
$T_{EPGA}(K:B)$	0.0347*	160.43	960	1.000	724.02

PRIORITIZATION BASED ON BOTH APPROACHES:

$$\begin{aligned}
 H(KWHUSE) = & T(KWHUSE:ENRATT) + T_{ENRATT}(KWHUSE:PERCEPT) \\
 & + T_{ENRATT,PERCEPT}(KWHUSE:GENATT) \\
 & + T_{ENRATT,PERCEPT,GENATT}(KWHUSE:AWARE) \\
 & + T_{ENRATT,PERCEPT,GENATT,AWARE}(KWHUSE:BEHAVE) \\
 & + H_{ENRATT,PERCEPT,GENATT,AWARE,BEHAVE}(KWHUSE)
 \end{aligned}$$

Research Question #4. This research question requires an analysis of the dependency (Conant 1982) in the system of variables. Dependency analysis can be performed using results of the SYSENT program, which can treat an arbitrary number of variables and classes. First, a reduced (four-variable) will be analyzed. Next, the total system of segmentation variables, in its original form will be assessed.

To construct the simplified model, the distinguished segmentation variable, energy consumption (KWHUSE) was included along with three top variables in the prioritization step. These variables were energy related attitudes, ENRATT, energy related perceptions, PERCEPT, and general attitudes, GENATT.

Dependency Analysis of the Four-Variable System of KWHUSE, ENRATT, PERCEPT, and GENATT. Table XX presents the dependency analysis results of the four-variable system of KWHUSE, ENRATT, PERCEPT, and GENATT.

In Table XX, column 1 gives the variable under consideration, column 2 gives the level of dependency, e.g., level 1 (n=1) tests one variable dependencies, level 2 (n=2) tests two-variable dependencies, etc. Column 3 gives the actual variables whose dependency to the variable in column 1 is being assessed.

TABLE XX
DEPENDENCY ANALYSIS OF THE FOUR-VARIABLE
SYSTEM OF GENATT, ENRATT, PERCEPT, AND
KWHUSE

(1)	(2)	(3)	(4)	(5)	(6)	(7)
i	n	S_i	$T(i:S_i)$	$T_n/H(i)$	%	$T(n)-T(n-1)$
K	1	E	0.0076	0.0053	12.60	0.0076
	2	EP	0.0239	0.0168	39.64	0.0163
	3	EPG	0.0603	0.0423	100.00	0.0364
E	1	G	0.0078	0.0035	11.22	0.0078
	2	GK	0.0286	0.0128	41.15	0.0208
	3	GKP	0.0695	0.0310	100.00	0.0409
P	1	G	0.0434	0.0317	44.83	0.0434
	2	GE	0.0586	0.0428	60.54	0.0125
	3	GEK	0.0968	0.0707	100.00	0.0382
G	1	P	0.0434	0.0279	44.51	0.0434
	2	PE	0.0611	0.0392	62.67	0.0177
	3	PEK	0.0975	0.0626	100.00	0.0364

Column 4 presents the bivariate transmissions between variables in column 1 and column 3 (variables in column 3 are taken as one aggregate variable). Column 5 gives a normalized measure of dependency (i.e., normalized by the entropy in the variable in column 1). Column 6 gives the percent of total dependency accounted for by level n. Finally, column 7 gives the amount dependency attributed to each additional level.

Using the information in Table XX and a threshold level of 80%, the dependency structures for each variable

are presented in Figure 13. Each graph includes the variables selected with respect to the threshold level for one of the variable (Figure 13, a thru d). Figure 13 (e) shows the combined structure constructed from the individual dependencies.

As mentioned in the methodology chapter, each variable selected carries the assumption that all variables higher in the ranking are present in the final structure and contribute indirectly to its measure of dependence (i.e., partial transmission). For example, in Figure 13 (a), GENATT contributes 0.0364 to KWHUSE if the relation KWHUSE <--> ENRATT <--> PERCEPT is present, otherwise, it contributes 0.0014 on its own.

Referring to Figure 13, the structure is constructed as KEP/EKP/EPG/GEP. However, KEP and EKP are the same and EPG and GEP are also identical. Therefore, the structure becomes KEP/EPG. This structure is presented Figure 13 (f).

The interpretation of this structure may be stated as follows. The strongest dependency exists between general attitudes, GENATT, and energy related perceptions, PERCEPT (0.0434). GENATT also has a strong effect of energy related attitudes (0.0078). Energy related attitudes, ENRATT, and perceptions, PERCEPT, in turn are directly interdependent with electricity consumption, KWHUSE.

Dependency Analysis of the Six-Variable System of KWHUSE, ENRATT, PERCEPT, GENATT, AWARE, and BEHAVE. The six variable system of segmentation variables can be analyzed through Conant's method without any alterations in the number of variables or classes. Table XXI shows the dependency analysis results.

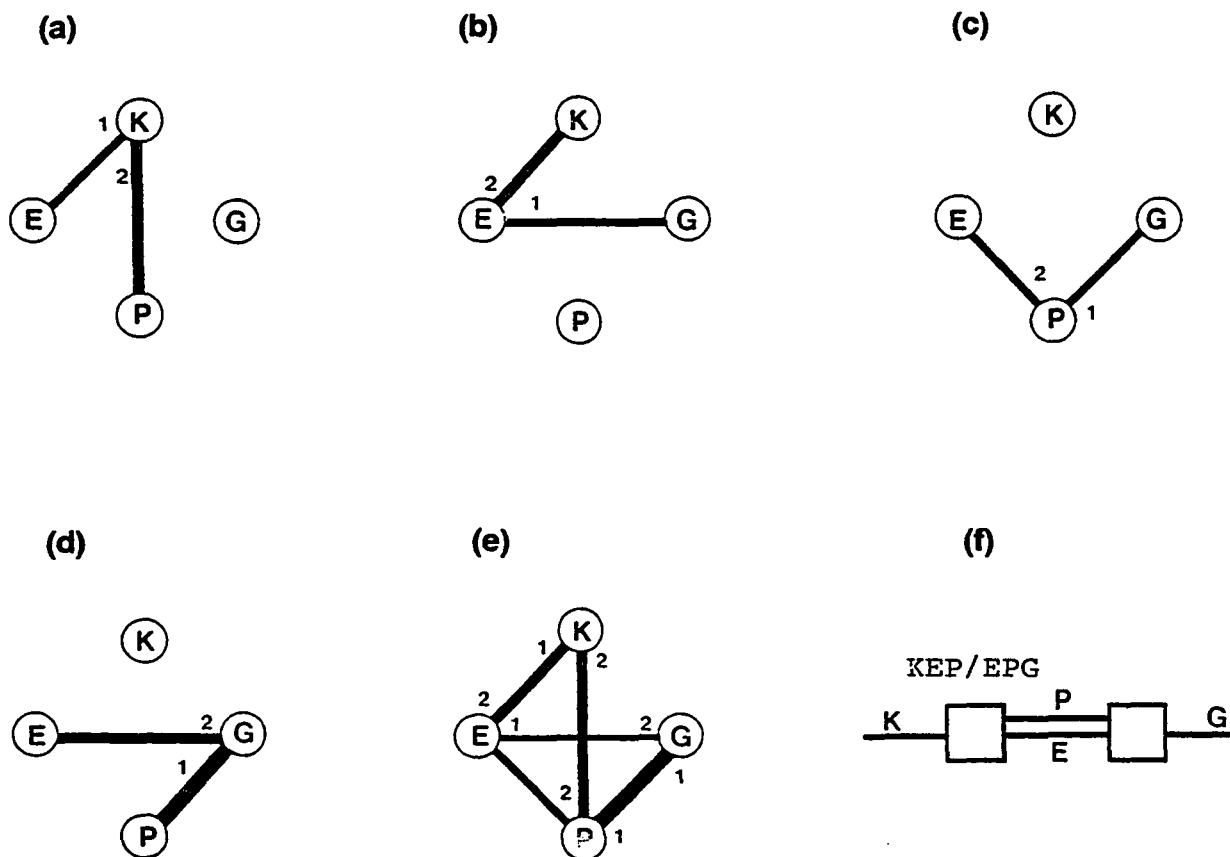


Figure 13. Graphical representation of the dependency analysis for the four-variable system of GENATT, ENRATT, PERCEPT, and KWHUSE.

The choice of relations to be retained (based on an 80% threshold) and the final structure are presented in Figure 14 (a) thru (f) and Figure 14 (g) respectively. The structure obtained can be presented in the spectral analysis format. The structure initially has the form of GPEK/EGKP/PGEK/ABK/BA/KEPGA. However, GPEK, EGKP, and PGEK are not only identically similar, but, also, embedded in KEPGA. Similarly, BA is embedded in ABK. Therefore, the structure KEPGA/ABK emerges. This structure is presented in Figure 14 (h).

The graph in Figure 14 (g) is interpreted as variables GENATT, ENRATT, PERCEPT, and KWHUSE being interdependent with varying (but close) strength of relationships. However, the variable AWARE is only related to KWHUSE (significantly) and there is a large amount of interdependence between the variables AWARE and BEHAVE. It is interesting to notice the BEHAVE which was analytically eliminated in the prioritization process (Research Question #3, Phase II) seems to have little significant direct relationship with the remaining variables.

TABLE XXI
DEPENDENCY ANALYSIS OF THE SIX-VARIABLE
SYSTEM OF SEGMENTATION VARIABLES

(1)	(2)	(3)	(4)	(5)	(6)	(7)
i	n	S _i	T(i:S _i)	T _n /H(i)	%	T(n)-T(n-1)
G	1	P	0.0434	0.0279	23.39	0.0434
	2	PE	0.0611	0.0392	32.86	0.0177
	3	PEK	0.0975	0.0626	52.47	0.0364
	4	PEKA	0.1550	0.0995	83.40	0.0575
	5	PEKAB	0.1858	0.1193	100.00	0.0308
E	1	G	0.0078	0.0035	4.63	0.0078
	2	GK	0.0286	0.0128	16.93	0.0208
	3	GKP	0.0695	0.0310	41.01	0.0409
	4	GKPA	0.1374	0.0613	81.09	0.0679
	5	GKPAB	0.1695	0.0756	100.00	0.0377
P	1	G	0.0434	0.0317	23.29	0.0434
	2	GE	0.0586	0.0428	31.45	0.0152
	3	GEK	0.0968	0.0707	51.95	0.0382
	4	GEKA	0.1527	0.1116	82.00	0.0559
	5	GEKAB	0.1863	0.1361	100.00	0.0336
A	1	B	0.3935	0.3884	78.40	0.3935
	2	BK	0.3989	0.3937	79.47	0.0054
	3	BKE	0.4138	0.4085	82.46	0.0149
	4	BKEP	0.4442	0.4385	88.51	0.0304
	5	BKEPG	0.5019	0.4954	100.00	0.0577
B	1	A	0.3935	0.6361	86.16	0.3935
	2	AK	0.4004	0.6473	87.67	0.0069
	3	AKE	0.4090	0.6612	89.56	0.0086
	4	AKEP	0.4259	0.6885	93.26	0.0169
	5	AKEPB	0.4567	0.7383	100.00	0.0308
K	1	E	0.0076	0.0053	4.82	0.0076
	2	EP	0.0239	0.0168	15.29	0.0163
	3	EPG	0.0603	0.0423	38.49	0.0364
	4	EPGA	0.1220	0.0856	77.89	0.0617
	5	EPGAB	0.1567	0.1099	100.00	0.0347

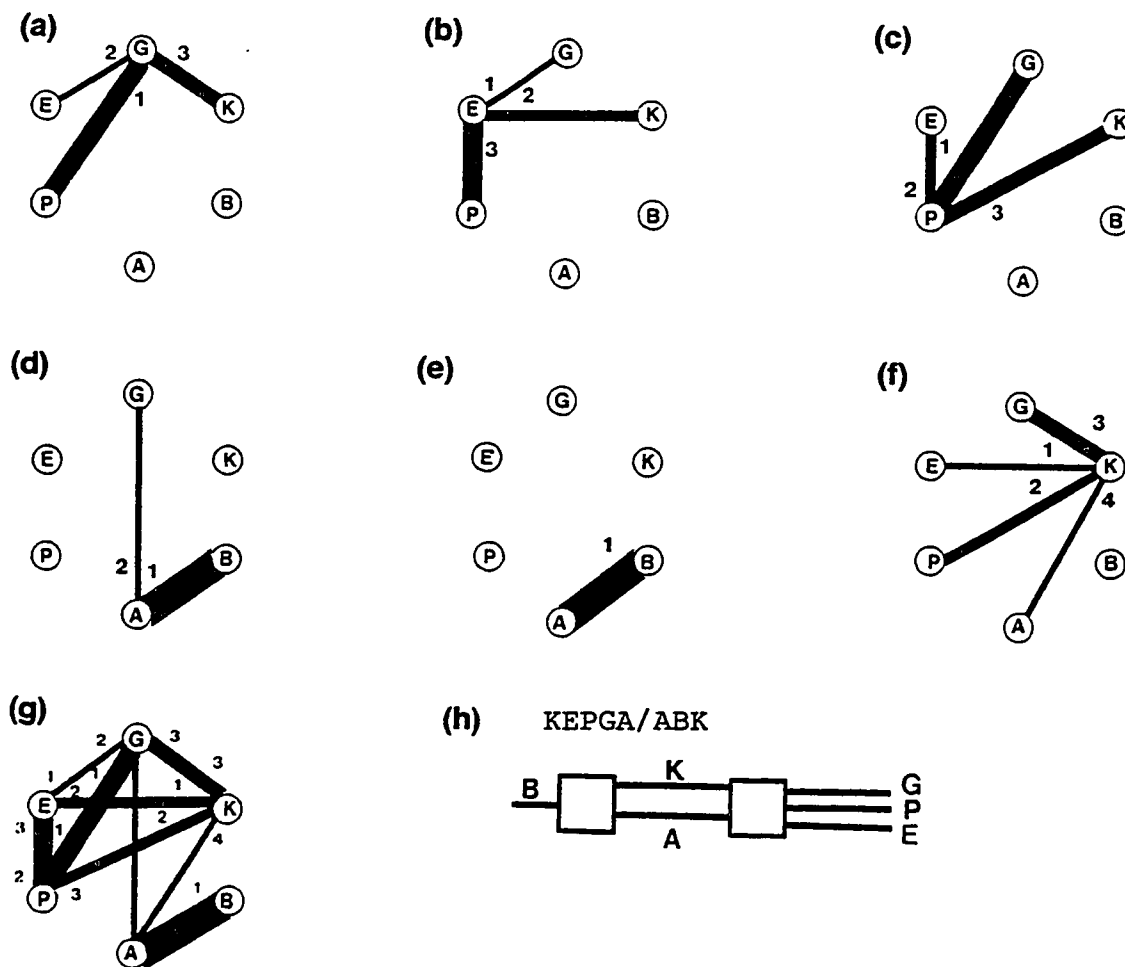


Figure 14. Graphical representation of the dependency structures for the six-variable system of segmentation variables.

Analytical Identification of the Distinguished Segmentation Variable

Research Question #1. This research question is intended to assess the predictability of each segmentation variable from a set descriptor variables analytically. For the current analyses, each candidate segmentation variable

is examined for predictability by the other segmentation variables. Table XXII presents the results obtained from the Logit procedure (which is an option within the LOGLINEAR procedure in SPSS^X).

In Table XXII, the logit analysis for each variable, column 1, produces the goodness-of-fit likelihood ratio chi-square, column 2, and its associated degrees of freedom, and risk level, columns 3 and 4, respectively. Column 5, 6 and seven, as mentioned earlier, list the measures of entropy, concentration, and the ranking of predictions, respectively. It is evident, based on the ranking, that BEHAVE is the distinguished variable, followed by AWARE, PERCEPT, KWHUSE, and finally, ENRATT.

Again, the information theory model, in table XXII, measures (in column 2) the uncertainty in each variable (as represented by its entropy), the uncertainty of that variable, given the descriptor variables, columns 3 and the amount and the percent of uncertainty removed by the descriptor variables, and the ranking of prediction are given in columns 4 and 5, respectively .

Therefore, the energy related behavior, BEHAVE, is designated as the distinguished variable. In continuing the analyses, therefore, BEHAVE will be treated as the distinguished variable.

TABLE XXII
PREDICTABILITY OF CANDIDATES FOR THE
DISTINGUISHED SEGMENTATION VARIABLE

LOG-LINEAR APPROACH (THE LOGIT MODEL)

(1) Candidate (i)	(2) L^2_{LL}	(3) df	(4) p	(5) <u>Ana. of Disper.</u> Entropy	(6) Concent	(7) Rank
ENRATT	683.08	1864	1	0.0122	0.0097	5
PERCEPT	580.05	1755	1	0.0445	0.0371	3
AWARE	445.30	1568	1	0.4004	0.3716	2
BEHAVE	247.14	1183	1	0.6519	0.6306	1
KWHUSE	623.47	1864	1	0.0153	0.0099	4

INFORMATION THEORETIC APPROACH

(1) Candidate (i)	(2) Uncert. of seg. Var. $H(i)$	(3) Uncert of seg.Var. given others $H_{\cdot}(i)$	(4) Amount (%) Uncert. Predicted	(5) Rank
ENRATT	2.2413	2.0663	0.1750 (7.81%)	5
PERCEPT	1.3687	1.1824	0.1863 (13.61%)	3
AWARE	1.0131	0.5111	0.5020 (49.55%)	2
BEHAVE	0.6186	0.1618	0.4568 (73.84%)	1
KWHUSE	1.4256	1.2689	0.1567 (10.99%)	4

Research Question #2. As the energy related behavior has only two classes, the second research question in Phase II reduces to the results obtained in question #1 of this phase for the variable BEHAVE. These results are reproduced in Table XXIII.

TABLE XXIII

PREDICTABILITY OF A PARTICULAR CLASS VS. ALL
OTHER CLASSES OF ENERGY RELATED BEHAVIOR

<u>LOG-LINEAR APPROACH (THE LOGIT MODEL)</u>						
(1) Reclassi- fication Scheme for KWHUSE	(2) L^2_{LL}	(3) df	(4) p	(5) <u>Ana. of Disper.</u> Entropy	(6) Concent	(7)
Class 1						
Class 2	247.14	1183	1	0.6519	0.6306	
<u>INFORMATION THEORETIC APPROACH</u>						
(1) Reclassi- fication Scheme for KWHUSE	(2) Uncert. of seg. Var. H(K)	(3) Uncert of seg.Var. given others H _. (K)	(4) Amount (%) Uncert. Predicted	(5)		
Class 1						
Class 2	0.6186	0.1618	0.2568 (73.84%)			

Descriptions of various columns are given in the previous part, in relation to Table XVIII, which is similar

to the above table. The high values for the analysis of dispersion (entropy measure = 0.6519, and concentration = 0.6306) for the log-linear model, and the amount of entropy in behave represented by the other variables, indicate that energy related behavior is highly predictable with the knowledge of the other segmentation variables.

Research Question #3. As expected, both the log-linear approach and the information theoretic approach give the same prioritized set of descriptor variables for BEHAVE. These results are given in Table XXIV.

Therefore, variables AWARE, KWHUSE, ENRATT, PRECEPT, GENATT, and BEHAVE form the sequence of priority of contribution to the reduction of uncertainty in BEHAVE.

Research Question #4. Dependency analysis (similar to other information theoretic structural modeling tools) is a symmetric technique (i.e., it investigates multiple variable simultaneously). The "optimal" dependency structure (subject to the threshold level) for the six-variable system of GENATT, ENRATT, PERCEPT, AWARE, BEHAVE, and KWHUSE will be the same regardless of which variable is considered as the dependent variable by the investigator. Therefore, the six-variable dependency structure given in connection with KWHUSE as the distinguished variable (see Table XXI and Figure 14), will be the same here, and will not be repeated.

TABLE XXIV
PRIORITIZATION OF THE EFFECT OF THE OTHER
SEGMENTATION VARIABLES ON ENERGY RELATED
BEHAVIOR (BEHAVE)

LOG-LINEAR APPROACH

Log-Linear Model	L^2_{LL}	df	p
(B)(G)	11.76	3	0.008
(B)(E)	16.41	4	0.003
(B)(P)	13.09	3	0.004
(B)(A)	1819.29*	2	0
(B)(K)	9.91	4	0.042
(B)(AG)	1833.95	11	0
(B)(AE)	1834.28	14	0
(B)(AP)	1826.13	11	0
(B)(AK)	1851.30*	14	0
(B)(AKG)	1887.27	61	0
(B)(AKE)	1891.07*	74	0
(B)(AKP)	1878.44	60	0
(B)(AKEG)	1956.46	299	1
(B)(AKEP)	1969.38*	299	1
(B)(AKEPG)	2111.75*	1199	1

INFORMATION THEORETIC APPROACH

Model	T	L^2_{IT}	df	p	Cum L^2_{IT}
T(B:G)	0.0025	11.55	3	0.008	
T(B:E)	0.0035	16.18	4	0.003	
T(B:P)	0.0028	12.95	3	0.004	
T(B:A)	0.3935*	1819.27	2	0	1819.27
T(B:K)	0.0021	9.71	4	0.042	
T _A (B:G)	0.0032	14.80	9	0	
T _A (B:E)	0.0032	14.80	12	0	
T _A (B:P)	0.0015	6.94	9	0	
T _A (B:K)	0.0069*	31.90	12	0	1851.17
T _{AK} (B:G)	0.0078	36.06	45	0	
T _{AK} (B:E)	0.0086*	39.76	60	0	1890.73
T _{AK} (B:P)	0.0059	27.28	45	0	
T _{AKE} (B:G)	0.0141	65.19	225	1	
T _{AKE} (B:P)	0.0169*	78.13	225	1	1970.06
T _(AKEP) (B:G)	0.1018*	142.40	900	1	2111.46

PRIORITIZATION BASED ON BOTH APPROACHES:

$$\begin{aligned}
 H(\text{BEHAVE}) = & T(\text{BEHAVE:AWARE}) + T_{\text{AWARE}}(\text{BEHAVE:KWHUSE}) \\
 & + T_{\text{AWARE,KWHUSE}}(\text{BEHAVE:ENRATT}) \\
 & + T_{\text{AWARE,KWHUSE,ENRATT}}(\text{BEHAVE:PERCEPT}) \\
 & + T_{\text{AWARE,KWHUSE,ENRATT,PERCEPT}}(\text{BEHAVE:GENATT}) \\
 & - H_{\text{AWARE,KWHUSE,ENRATT,PERCEPT,GENATT}}(\text{BEHAVE})
 \end{aligned}$$

Next, the three variables in the prioritized set of descriptor variables were selected for structure analysis using Conant's method. The variables are AWARE (A), KWHUSE (K), and ENRATT (E) along with the distinguished variable BEHAVE (B).

This four-variable system (i.e., BEHAVE, AWARE, KWHUSE, ENRATT), however, is different from that analyzed in connection with KWHUSE. This is due to the different orders of prioritization obtained for KWHUSE (as the distinguished variable) versus BEHAVE.

Dependency Analysis of the Four-Variable System of BEHAVE, AWARE, KWHUSE, AND ENRATT. Table XXV presents the dependency analysis results of the four-variable system of KWHUSE, ENRATT, PERCEPT, and GENATT.

TABLE XXV

DEPENDENCY ANALYSIS OF THE FOUR-VARIABLE
SYSTEM OF BEHAVE, AWARE, KWHUSE, AND ENRATT

(1)	(2)	(3)	(4)	(5)	(6)	(7)
i	n	S_i	$T(i:S_i)$	$T_n/H(i)$	%	$T(n)-T(n-1)$
B	1	A	0.3935	0.6361	96.21	0.39325
	2	AK	0.4004	0.6473	97.90	0.0069
	3	AKE	0.4090	0.6612	100.00	0.0054
A	1	B	0.3935	0.3884	95.09	0.3935
	2	BK	0.3989	0.3937	96.40	0.0054
	3	BKE	0.4138	0.4085	100.00	0.0149
K	1	E	0.0076	0.0053	26.95	0.0016
	2	EA	0.0159	0.0112	56.38	0.0015
	3	EAB	0.0282	0.0198	100.00	0.0021
E	1	K	0.0076	0.0034	26.12	0.0076
	2	KA	0.0205	0.0092	70.45	0.0129
	3	KAB	0.0291	0.0130	100.00	0.0086

Using the threshold level of 80%, the dependency structures with respect to each variable are given in Figure 15 (a thru d). The total dependency structure is presented in Figure 15 (e). The structure in Figure 15 (e) is shown in it alternative format in Figure 15 (f).

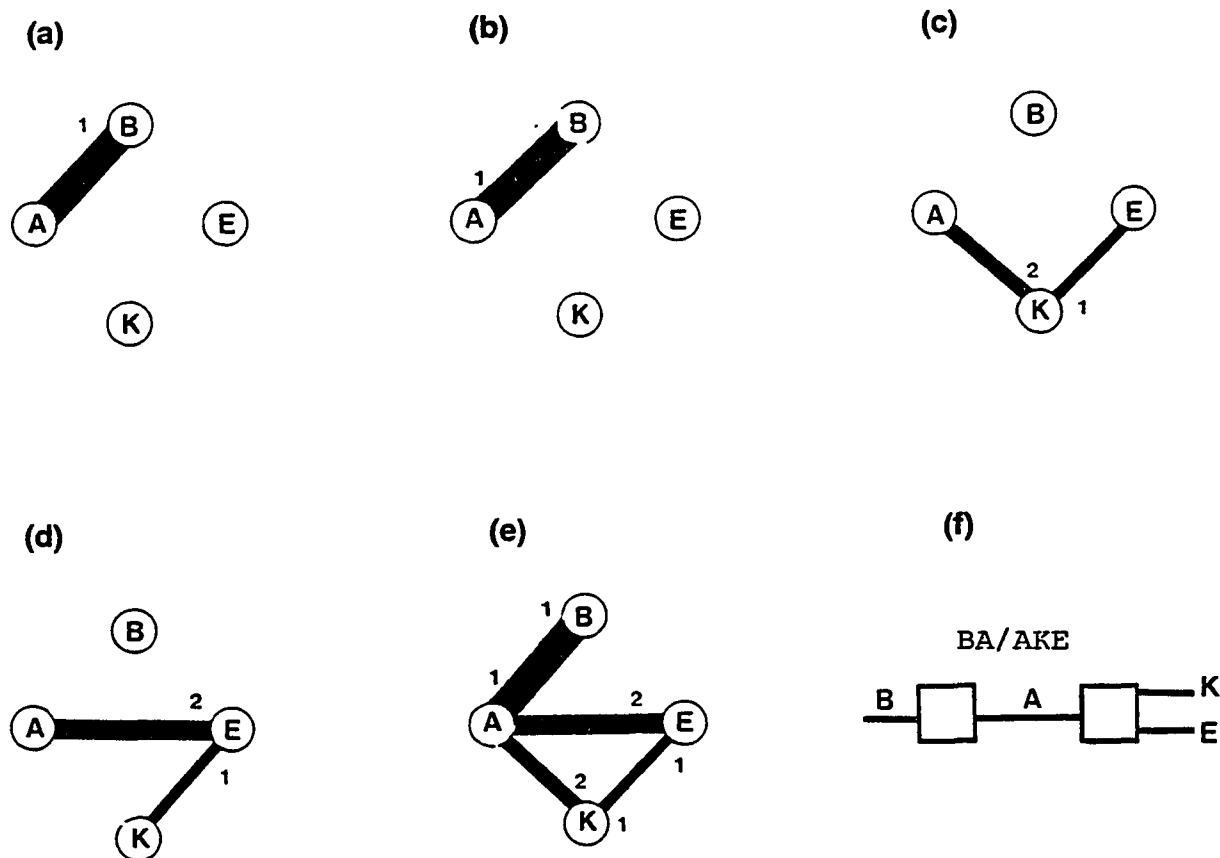


Figure 15. Graphical representation for the dependency structures for the four-variable system of BEHAVE, AWARE, KWHUSE, and ENRATT.

After removing the embedded and identical subsystems from BA/AKE/KEA/EAK, the structure BA/AKE (Figure 15 (f)) emerges. This structure can be interpreted as follows. Energy related behavior, BEHAVE, directly interact with energy related awareness, AWARE, which has a direct effect on electricity consumption, KWHUSE, and energy related

attitudes, ENRATT. The latter three variables have mutual effect on each other.

DETERMINATION AND ANALYSIS OF THE DISTINGUISHED SEGMENTATION VARIABLE USING EXOGENOUS DESCRIPTOR VARIABLES

The previous section conducted Phase II of segment congruence analysis using endogenous variables (i.e., other segmentation variables), as descriptor variable, to identify the distinguished segmentation variable and analyze it through structure analysis techniques. This section introduces a set of exogenous variables (or clusterings) to be used as descriptor variables. Then the analysis in Phase II will be repeated using these new descriptor variables.

Table XI in chapter III presents these descriptor variables, and Table X lists the variables selected from the PNWRES data (see Appendix IV for a complete list of variables). Also, in Chapter III, the procedures used to obtain and cluster all of the variables used in this study are described. Table XXVI reproduces the listing of the exogenous variables for ease of reference.

The analysis will proceed with the evaluation of the candidate segmentation variables, obtained in section 5.1. through the above descriptor variables. These candidate variables are ENRATT, PERCEPT, AWARE, BEHAVE, and KWHUSE.

TABLE XXVI
LIST OF THE EXOGENOUS DESCRIPTOR VARIABLES

Variable (Abbreviation)	Description	# of Classes
CLIMGEO (C)	Climatic/Geographic Environment	3
TYPDWEL (T)	Type of the Dwelling Unit	4
RENTOWN (R)	The Dwelling is Rental/Owned	3
DEMOG (D)	Demographic Characteristics	4
INSUL (I)	Level of Insulation	3

Research Question #1. Table XXVII shows the results of five logit models, each using one of the candidates for the distinguished variable as the dependent variable against the exogenous descriptor variables. The information theoretic analysis, also presented in Table XXVII, calculates the amount of uncertainty in each candidate variable reduced by the descriptor variables.

It is seen that both methods result in the same distinguished variable (i.e., KWHUSE). The two methods, as expected, give identical ranking of the candidates for the distinguished variable, as well.

TABLE XXVII

PREDICTABILITY OF CANDIDATES FOR THE
DISTINGUISHED SEGMENTATION VARIABLE USING
EXOGENOUS DESCRIPTOR VARIABLES

<u>LOG-LINEAR APPROACH (THE LOGIT MODEL)</u>						
(1) Candidate (i)	(2) L^2_{LL}	(3) df	(4) p	(5) <u>Ana. of Disper.</u> Entropy	(6) Concent	(7) Rank
ENRATT	627.89	1680	1	0.0209	0.0171	4
PERCEPT	400.12	1260	1	0.0110	0.0103	5
AWARE	330.90	840	1	0.0336	0.0358	3
BEHAVE	163.81	420	1	0.0504	0.0482	2
KWHUSE	408.95	1640	1	0.0632	0.0392	1

<u>INFORMATION THEORETIC APPROACH</u>				
(1) Candidate (i)	(2) Uncert. of seg. Var. H(i)	(3) Uncert of seg.Var. given others H _. (i)	(4) Amount (%) Uncert. Predicted	(5) Rank
ENRATT	2.2447	2.0625	0.1822 (8.12%)	4
PERCEPT	1.3581	1.2486	0.1095 (8.06%)	5
AWARE	1.0354	0.9284	0.1070 (10.33%)	3
BEHAVE	0.6436	0.5760	0.0676 (10.50%)	2
KWHUSE	1.4171	1.2304	0.1867 (13.18%)	1

Having identified the distinguished segmentation variable KWHUSE, the remaining analyses concentrate on

assessing the relationships and the structure of the system composed of this variable and the descriptor variables. First, the question of predictability will be pursued further in question #2. Next, the descriptor variables will be prioritized in question #3. Finally, the structure of relationships will be assessed in questions #4 and #5.

Research Question #2. The distinguished variable (KWHUSE) was reclassified based on three collapsing schemes (i.e., the same three groupings as used in section 5.2). First, the low-consumption group was aggregated (i.e., classes 1 and 2) as class #1, and the remaining classes were aggregated (i.e., classes 3, 4, and 5) as class #2. Then, the medium-consumption class (i.e., class #3) was dichotomized against all other classes combined. Finally, the high-consumption group (i.e., classes 4 and 5) were dichotomized against all other classes.

Then, both the log-linear approach (i.e., logit models) and the information theoretic approach were used to assess the differences in the three grouping schemes. The results are presented in Table XXVIII.

It is evident that the high-consumption/other-groups scheme has the highest predictability measures followed by the low-consumption/others, and medium-consumption/others, as second and third best represented schemes, respectively.

TABLE XXVIII

PREDICTABILITY OF A PARTICULAR CLASS VS. ALL
OTHER CLASSES OF THE DISTINGUISHED VARIABLE,
(KWHUSE)

LOG-LINEAR APPROACH

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Reclassi- fication Scheme for KWHUSE	L^2_{LL}	df	p	<u>Ana. of Disper.</u> <u>Entropy</u> <u>Concent</u>		
Class 1&2 = 1 Others = 2	163.79	420	1	0.0466	0.0560	
Class 3 = 1 Others = 2	193.55	420	1	0.0176	0.0235	
Class 4&5 = 1 Others = 2	145.07	420	1	0.0945	0.0717	

INFORMATION THEORETIC APPROACH

(1)	(2)	(3)	(4)	(5)
Reclassi- fication Scheme for KWHUSE	Uncert. of seg. Var. H(K)	Uncert of seg.Var. given others $H_o(K)$	Amount (%) Uncert. Predicted	
Class 1&2 = 1 Others = 2	0.8731	0.7934	9.13	
Class 3 = 1 Others = 2	0.9700	0.9070	6.50	
Class 4&5 = 1 Others = 2	0.4848	0.4045	16.56	

This, basically, indicates that given the descriptor variables, the uncertainty with respect to high-consumption/others is less than other groupings.

Research Question #3. Table XXIX presents the log-linear approach and the information theoretic approach to prioritization of contributions of the descriptor variables to the distinguished segmentation variable, KWHUSE. At the bottom of the table, this prioritization is expressed in information theoretic terms.

Here, the descriptor variables DEMOG, TYPDWEL, CLIMGEO, INSUL, and RENTOWN contribute to the representation of distinguished variable. Their priority is in the sequence given above. Therefore, the three most significant descriptor variables needed for research question #4 are DEMOG, TYPDWEL, and CLIMGEO.

TABLE XXIX
PRIORITIZED EFFECT OF THE EXOGENOUS DES-
CRIPTOR VARIABLES ON ENERGY CONSUMPTION

LOG-LINEAR APPROACH

Log-Linear Model	L^2_{LL}	df	p
(K)(C)	56.87	8	= 0
(K)(T)	175.22	12	= 0
(K)(R)	46.78	8	= 0
(K)(D)	182.35*	12	= 0
(K)(I)	28.75	8	= 0
(K)(DC)	260.81	44	= 0
(K)(DT)	313.10*	63	= 0
(K)(DR)	247.06	44	= 0
(K)(DI)	217.56	44	= 0
(K)(DTC)	442.06*	211	= 0
(K)(DTR)	399.17	211	= 0
(K)(DTI)	386.90	210	= 0
(K)(DTCR)	565.32	595	= 0
(K)(EPGB)	587.93*	595	= 0
(K)(EPGAB)	786.05*	1747	= 0

INFORMATION THEORETIC APPROACH

Model	T	L^2_{IT}	df	p	Cum L^2_{IT}
T(K:C)	0.0135	56.84	8	= 0	
T(K:T)	0.0416	175.14	12	= 0	
T(K:R)	0.0111	46.73	8	= 0	
T(K:D)	0.0433*	182.30	12	= 0	182.30
T(K:I)	0.0068	28.63	8	= 0	
$T_D(K:C)$	0.0186*	78.31	32	= 0	
$T_D(K:T)$	0.0311*	130.94	51	= 0	313.24
$T_D(K:R)$	0.0154	64.84	32	= 0	
$T_D(K:I)$	0.0080	33.68	32	= 0	
$T_{DT}(K:C)$	0.0286*	120.41	148	= 0	433.65
$T_{DT}(K:R)$	0.0168	70.73	148	= 0	
$T_{DT}(K:I)$	0.0150	63.15	148	= 0	
$T_{DTC}(K:R)$	0.0314*	132.20	384	= 0	
$T_{DTC}(K:I)$	0.0367*	154.51	384	= 0	588.16
$T(DTCI(K:R))$	0.0471*	198.30	1152	= 0	786.46

PRIORITIZATION BASED ON BOTH APPROACHES:

$$\begin{aligned}
 H(KWHUSE) = & T(KWHUSE:DEMOG) + T_{DEMOG}(KWHUSE:TYPDWEL) \\
 & + T_{DEMOG,TYPDWEL}(KWHUSE:CLIMGEO) \\
 & + T_{DEMOG,TYPDWEL,CLIMGEO}(KWHUSE:INSUL) \\
 & + T_{DEMOG,TYPDWEL,CLIMGEO,INSUL}(KWHUSE:RENTOWN) \\
 & + H_{DEMOG,TYPDWEL,CLIMGEO,INSUL,RENTOWN}(KWHUSE)
 \end{aligned}$$

Research Question #4. Dependency Analysis of the Four-Variable System of KWHUSE, DEMOG, TYPDWEL, and CLIMGEO. The results of dependency analysis of KWHUSE, DEMOG, TYPDWEL, and CLIMGEO are given in Table XXX.

TABLE XXX
DEPENDENCY ANALYSIS OF THE FOUR-VARIABLE
SYSTEM OF KWHUSE, DEMOG, TYPDWEL, AND
CLIMGEO

(1)	(2)	(3)	(4)	(5)	(6)	(7)
i	n	S _i	T(i:S _i)	T _n /H(i)	%	T(n)-T(n-1)
K	1	D	0.0433	0.0306	42.04	0.0433
	2	DT	0.0744	0.0525	72.23	0.0311
	3	DTC	0.1030	0.0727	100.00	0.0286
D	1	T	0.0762	0.0387	58.57	0.0762
	2	TK	0.1090	0.0553	83.78	0.0328
	3	TKC	0.1301	0.0661	100.00	0.0211
T	1	D	0.0762	0.0835	62.36	0.0762
	2	DK	0.1073	0.1175	87.81	0.0311
	3	DKC	0.1222	0.1339	100.00	0.0149
C	1	K	0.0135	0.0085	32.93	0.0135
	2	KD	0.0261	0.0165	63.66	0.0126
	3	KDT	0.0410	0.0259	100.00	0.0149

Again, assuming 80% threshold, the graphs of dependencies were obtained for each variable. These graphs along with the total graph are given in Figure 16 (a thru e). Also, the alternative representation of the total graph is given in Figure 16 (f).

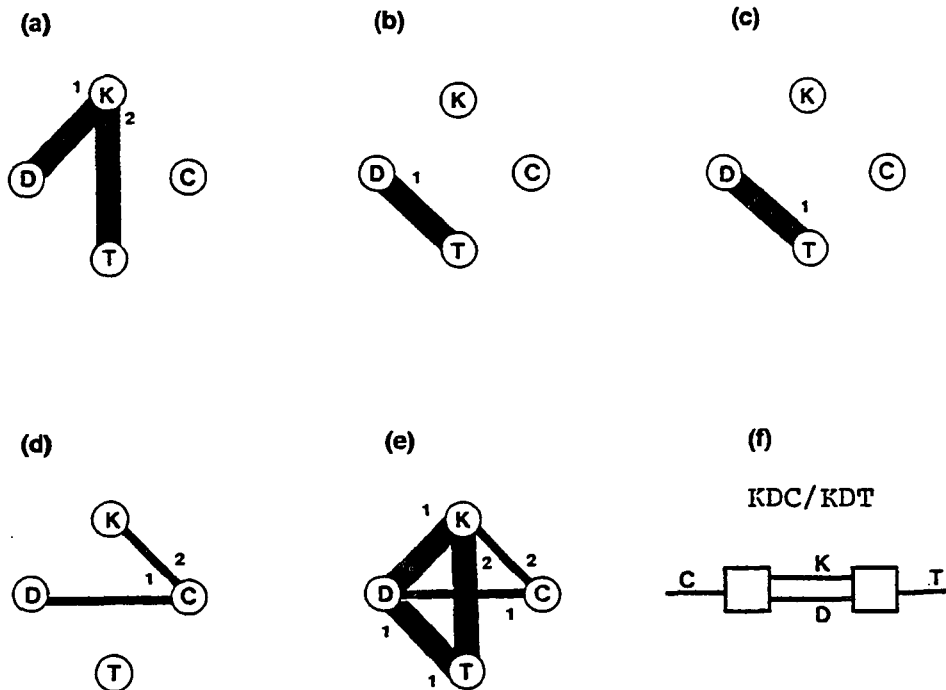


Figure 16. Graphical representation of the dependency structures for the four-variable system of KWHUSE, DEMOG, TYPDWEL, and CLIMGEO.

The emerging structural model based on Conant's dependency analysis is KDC/KDT. This structure implies that electricity consumption, KWHUSE, is directly affected by the type of dwelling, TYPDWEL, and the geoclimatic environment, CLIMGEO. The demographic variable has direct relationships with KWHUSE, and CLIMGEO on the one hand, and with KWHUSE and TYPDWEL on the other.

Dependency Analysis of the Six-Variable System of KWHUSE, ENRATT, PERCEPT, GENATT, AWARE, and BEHAVE. Table XXXI shows the dependency analysis results.

TABLE XXXI

DEPENDENCY ANALYSIS OF THE DISTINGUISHED
VARIABLE AND THE EXOGENOUS DESCRIPTOR
VARIABLES

(1)	(2)	(3)	(4)	(5)	(6)	(7)
i	n	S_i	$T(i:S_i)$	$T_n/H(i)$	%	$T(n)-T(n-1)$
K	1	D	0.0433	0.0306	23.22	0.0433
	2	DT	0.0744	0.0525	39.83	0.0311
	3	DTC	0.1030	0.0727	55.16	0.0286
	4	DTCI	0.1397	0.0986	74.81	0.0367
	5	DTCIR	0.1868	0.1318	100.00	0.0471
C	1	K	0.0135	0.0085	11.69	0.0135
	2	KD	0.0261	0.0165	22.70	0.0126
	3	KDI	0.0438	0.0277	38.10	0.0177
	4	KDIT	0.0745	0.0470	64.65	0.0307
	5	KDITR	0.1151	0.0727	100.00	0.0406
T	1	D	0.0762	0.0835	36.72	0.0762
	2	DK	0.1073	0.1175	51.67	0.0311
	3	DKC	0.1492	0.1634	71.86	0.0419
	4	DKCI	0.1753	0.1920	84.43	0.0261
	5	DKCIR	0.2076	0.2274	100.00	0.0323
R	1	D	0.0217	0.0321	17.32	0.0217
	2	DI	0.0371	0.0549	29.63	0.0154
	3	DIK	0.0628	0.0929	50.14	0.0257
	4	DIKC	0.0930	0.1375	74.20	0.0302
	5	DIKCT	0.1253	0.1853	100.00	0.0323
D	1	T	0.0762	0.0387	34.52	0.0762
	2	TK	0.1090	0.0553	49.33	0.0328
	3	TKR	0.1328	0.0674	60.13	0.0238
	4	TKRC	0.1670	0.0848	75.65	0.0342
	5	TKRCI	0.2208	0.1121	100.00	0.0538
I	1	R	0.0148	0.0123	13.73	0.0148
	2	RD	0.0209	0.0174	19.42	0.0061
	3	RDK	0.0396	0.0330	36.83	0.0187
	4	RDKC	0.0715	0.0596	66.52	0.0319
	5	RDKCT	0.1075	0.0896	100.00	0.0360

The choice of relations to be retained (based on an 80% threshold) and the final structure are presented in Figure 17 (a) thru (f) and Figure 17 (g) respectively.

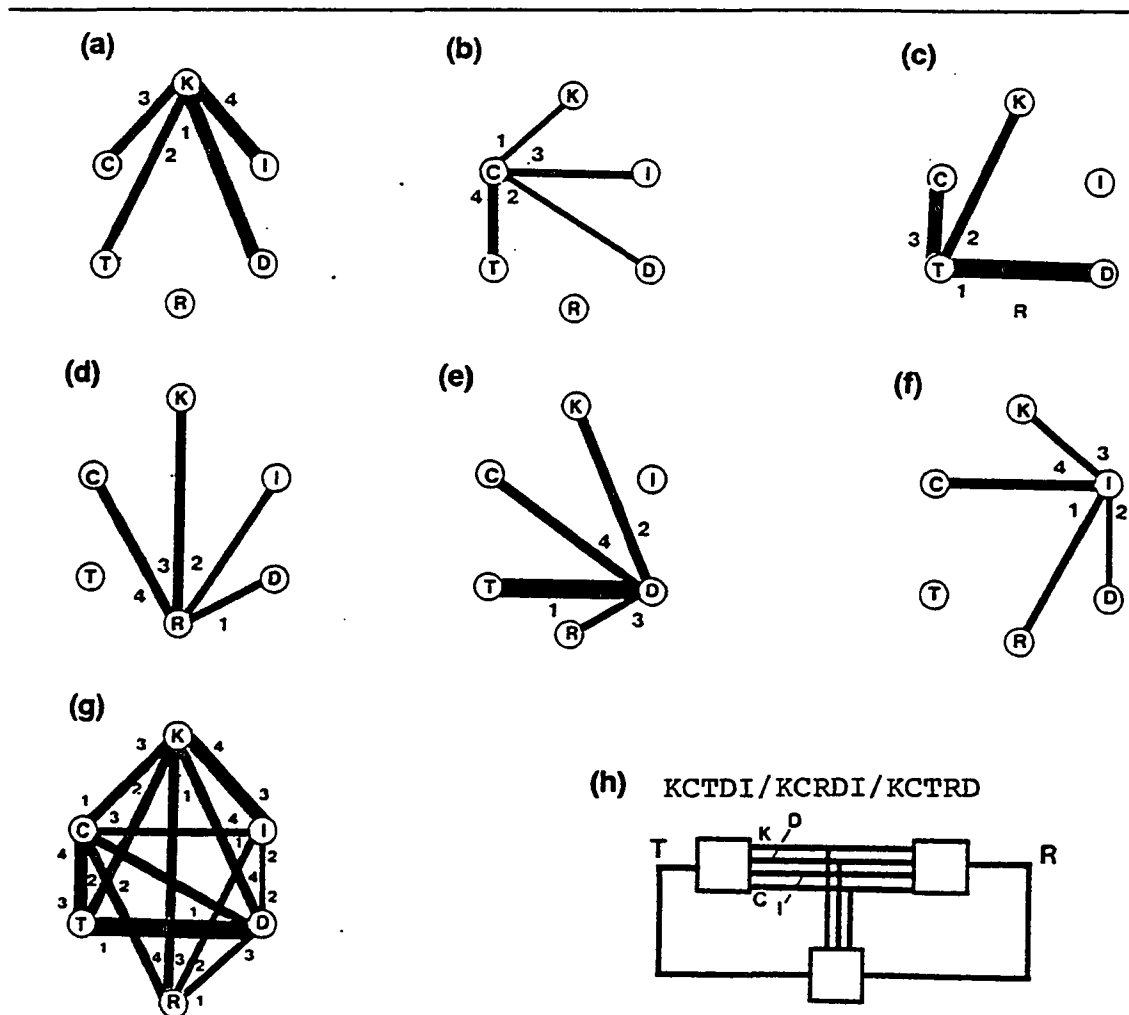


Figure 17. Graphical representation of the dependency structures for the system of distinguished and exogenous variables.

The dependency structure in Figure 17 (g) can be represented as KDTCI/CKDIT/TDKC/RDIKC/DTKRC/IRDKC. However, the first and the second terms in the above structure are identical. Also, TDKC (the third term) is embedded in the first two terms. Furthermore, the fourth and the sixth terms are identical. Therefore, the structure is rewritten, eliminating all but one of the identical terms, as KCTDI/KCRDI/KCTRD. This structure is given in Figure 17 (h).

SUMMARY OF ANALYSIS

This chapter used the PNWRES data to demonstrate that all of the proposed questions in segment congruence analysis can be addressed through the proposed methodology. The equivalent log-linear models were proposed to perform these analyses, where applicable and practical. Tables XXXII through XXXV present all of the results obtained throughout this chapter in a brief format.

It is clear that most of the proposed analyses can be performed using either method. However, all of the results were obtained through three main SYSENT runs plus some additional minor runs. The log-linear approach on the other hand required hundreds of runs. Additionally, the amounts of information and/or uncertainty contained in each model could not be obtained through the log-linear approach.

Therefore, it is proposed that the researchers use the information theoretic approach for exploratory analyses and identification of important variables. After these variables are identified, the algorithms common in log-linear computer packages (e.g., the proportional fitting method) can be used to obtain cell estimates and predictions in a disaggregate form.

TABLE XXXII

ANALYSIS OF THE MUTUAL ASSOCIATION AMONG
SEGMENTATION VARIABLES AND IDENTIFICATION
OF CANDIDATES FOR THE DISTINGUISHED
SEGMENTATION VARIABLE

Description	Information Theoretic Approach	Log-Linear/ Logit Approach
A) MUTUAL ASSOCIATION		
Existence	Yes	Yes
Amount	$T/T_{MAX}=0.0765$	-
B) CANDIDATES FOR THE DISTINGUISHED BASE		
Contribution to Mutual Association	AWARE	AWARE
Contribution to Joint Variability	(aware) BEHAVE	(aware) BEHAVE
Contribution to Uncertainty of System	ENRATT, (aware), (behave)	~
Number of Significant Dyadic Relations	PERCEPT	PERCEPT
JUDGMENT-BASED CANDIDATE	KWHUSE	KWHUSE

TABLE XXXIII
 ANALYSIS USING OTHER SEGMENTATION VARIABLES
 AS DESCRIPTOR VARIABLES
 ELECTRICITY CONSUMPTION AS THE DISTINGUISHED
 VARIABLE (A PRIORI)

Description	Information Theoretic Approach	Log-Linear/ Logit Approach
<hr/>		
PREDICTABILITY OF COMBINED CLASSES	(Ranking)	
Class 1&2/Others (Low)	(2)	(2)
Class 3/Others (Medium)	(3)	(3)
Class 4&5/Others (High)	(1)	(1)
<hr/>		
PRIORITIZATION OF DESCRIPTOR VARIABLES	(Priority)	
GENATT	(3)	(3)
ENRATT	(1)	(1)
PERCEPT	(2)	(2)
AWARE	(4)	(4)
BEHAVE	(5)	(5)
<hr/>		
STRUCTURE ANALYSIS		
The 4-variable system	EPK/PG	EPK/PG
The 6-variable system	KEPGA/ABK	KEPGA/ABK
<hr/>		

TABLE XXXIV
ANALYSIS USING OTHER SEGMENTATION VARIABLES
AS DESCRIPTOR VARIABLES
ANALYTICAL IDENTIFICATION AND ANALYSIS OF
THE DISTINGUISHED VARIABLE

Description	Information Theoretic Approach	Log-Linear/ Logit Approach
<hr/>		
THE DISTINGUISHED SEGMENTATION VARIABLE	(Ranking)	
ENRATT	(5)	(5)
PERCEPT	(3)	(3)
AWARE	(2)	(2)
BEHAVE *	(1)	(1)
KWHUSE	(4)	(4)
ANALYSIS OF THE DISTINGUISHED VARIABLE		
PREDICTABILITY OF COMBINED CLASSES	(Ranking)	
Class 1/2	(1)	(1)
PRIORITIZATION OF DESCRIPTOR VARIABLES		
	(Priority)	
GENATT	(5)	(5)
ENRATT	(3)	(3)
PERCEPT	(4)	(4)
AWARE	(1)	(1)
KWHUSE	(2)	(2)
STRUCTURE ANALYSIS		
The 4-variable system	KBA/EBK	KBA/EBK
The 6-variable system	KEPGA/ABK	KEPGA/ABK
* The variable selected as the distinguished segmentation base.		

TABLE XXXV
ANALYSIS USING EXOGENOUS DESCRIPTOR
VARIABLES

Description	Information Theoretic Approach	Log-Linear/ Logit Approach
<hr/>		
THE DISTINGUISHED SEGMENTATION VARIABLE		(Ranking)
ENRATT	(4)	(4)
PERCEPT	(5)	(5)
AWARE	(3)	(3)
BEHAVE	(2)	(2)
KWHUSE *	(1)	(1)
ANALYSIS OF THE DISTINGUISHED VARIABLE		
PREDICTABILITY OF COMBINED CLASSES		(Ranking)
Class 1&2/Others (Low)	(2)	(2)
Class 3/Others (Medium)	(3)	(3)
Class 4&5/Others (High)	(1)	(1)
PRIORITIZATION OF DESCRIPTOR VARIABLES		
		(Priority)
CLIMGEO	(3)	(3)
TYPDWEL	(2)	(2)
RENTOWN	(5)	(5)
DEMOG	(1)	(1)
INSUL	(4)	(4)
STRUCTURE ANALYSIS		
The 4-variable system	KT/TD/DC	KT/TD/DC
The 6-variable system	KCTDI/KCRDI/KCTRD	
* The variable selected as the distinguished segmentation base.		

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

This chapter presents the conclusions of the study as well as several suggestions for extensions and future directions in this area. First, a brief summary of the findings in this study is presented. Next, contributions of this study to (a) segment congruence analysis, (b) market segmentation in general, (c) log-linear modeling, and (d) information theoretic modeling are explored. Then, the limitations of this study are delineated. Finally, some unresolved problems and directions for further research are discussed.

FINDINGS OF THE STUDY

Multivariate segmentation studies are gaining in popularity due to the realization that, often, managerial judgment in selecting segmentation variables leads to incorrect definition and determination of market segments. One reason for this shortcoming is the dominant practice of identifying only one variable (e.g., usage rate) as the

segmentation variable, and a class of this variable (e.g., the heavy user) as the target segment.

In reality, however, the choice of a segmentation variable, as well as a target segment, are context dependent. In the electrical utility industry, for example, the electricity consumption is viewed as a segmentation variable with respect to, say, the strategy of identifying and retaining the heavy users. On the other hand, attitudes toward energy conservation is a segmentation variable if, for instance, the utility company plans to implement a conservation program.

Some segmentation variables are more "accessible" than others. Knowledge about the characteristics of the individuals associated with the more accessible segmentation variable may lead to a better understanding of the other, less accessible, variables.

Segment congruence analysis is composed of a family of techniques which deals with multiple variable segmentation studies. The classical segment congruence analysis entailed only a few questions with the primary purpose of assessing mutual association among segmentation variables, assessing the contribution of each segmentation variable to the total mutual association, and assessing the predictability of an a priori selected distinguished variable.

The present study extended segment congruence analysis by introducing several relevant research questions.

Furthermore, the information theoretic family of techniques was proposed as an approach useful for multivariate segmentation studies in the framework of segment congruence analysis. It was also demonstrated, both in the methodological and application sense, that the information theoretic approach is capable of addressing the old and the new questions in segment congruence analysis.

The results of this application, not only demonstrated the efficacy of the information theoretic and the log-linear approaches, but clearly demonstrated that the results of segment congruence analysis are meaningful to the practitioner as well as the researcher in market segmentation (In the next section of this chapter, contributions of this study will be discussed).

Tables XXXVI thru XXXIX outline all of the research questions in the two phases, the results of the PNWRES study application, and the practical and theoretical significance of these results. The assessment of mutual association and identification of candidate segmentation variables are given in Table XXXVI. The a priori selection of the distinguished variable and its analyses are given in Table XXXVII. The results of the "identification and analysis of the distinguished variable" using the analytical (as opposed to the a priori) approach, based on the endogenous and exogenous descriptor variables, are presented in Tables XXXVIII and XXXIX.

TABLE XXXVI

PRACTICAL AND/OR THEORETICAL SIGNIFICANCE OF
THE RESULTS OF SEGMENT CONGRUENCE ANALYSIS
TO THE PNWRES STUDY: PHASE I

Description	Result	Implications
<u>RESEARCH QUESTION #1</u> Are the segmentation variables mutually associated? If yes, how can this mutual association be measured?	YES $T/T_{MAX} = 0.0765$	Knowledge about other variables contributes to the understanding of a variable of particular interest.
<u>RESEARCH QUESTION #2</u> Which basis (or set of bases) makes the highest contribution toward the mutual association?	AWARE	Based on this criterion, energy related awareness is a candidate for the distinguished variable. (Method #1)
<u>RESEARCH QUESTION #3</u> Which variable makes the highest contribution toward the joint variability in the system of segmentation bases?	(aware) ¹ BEHAVE	Based on this criterion, energy related behavior is a candidate for the distinguished variable. (Method #2)
<u>RESEARCH QUESTION #4</u> Knowledge of which variable makes the highest contribution toward the reduction of overall uncertainty in the overall system of segmentation bases?	ENRATT	Based on this criterion, energy related attitudes is a candidate for the distinguished variable. (Method #3)
<u>RESEARCH QUESTION #5</u> Which variable has the most number of statistically significant dyadic interrelationships with other variables?	PERCEPT (aware) ² (behave) ²	Based on this criterion, energy related perceptions is a candidate for the distinguished variable. (Method #3)
JUDGMENT-BASES	KWHUSE	Key variable.

¹ BEHAVE, the second choice, was selected to increase the number of variables for this illustrative application.

² All three variables had 3 significant dyadic relations.

TABLE XXXVII

SIGNIFICANCE OF THE RESULTS OF SEGMENT
CONGRUENCE ANALYSIS TO THE PNWRES STUDY

PHASE II: A PRIORI DISTINGUISHED VARIABLE

Description	Result	Implications
<u>RESEARCH QUESTION #1</u> Which one of the candidate distinguished segmentation bases, identified in phase I, is best predicted once we have the knowledge of a set of relevant independent variables?	KWHUSE (10.99%)	Selected <u>a priori</u> .
<u>RESEARCH QUESTION #2</u> How much of the uncertainty in the distinguished variable is reduced when it is dichotomized into a class (or group of classes) versus all of the other classes?	11.18% (4&5 vs. others) 6.45% (1&2 vs. others) 6.10% (3 vs. others)	The classification of the electricity consumption variables to dichotomous variables (light, medium, or heavy users vs others) shows the heavy user segment is best predicted by the other segmentation variables.
<u>RESEARCH QUESTION #3</u> How can one prioritize the effect of the independent variables on the distinguished set?	1:ENRATT 2:PERCEPT 3:GENATT 4:AWARE 5:BEHAVE	In reducing the number of descriptor variables to be used in subsequent analyses, use the given order (ENRATT=Most-BEHAVE=least-important).
<u>RESEARCH QUESTION #4</u> What is the dependency structure in each class of the distinguished segmentation base?	EPK/PG KEPGA/ ABK	When four segmentation variables are used, indirect/direct effects in the simplest structure are as given. When all six segmentation variables are used, indirect/direct effects in the simplest structure are as given.

TABLE XXXVIII

SIGNIFICANCE OF THE RESULTS OF SEGMENT
CONGRUENCE ANALYSIS TO THE PNWRES STUDYPHASE II: OTHER SEGMENTATION VARIABLES AS
DESCRIPTOR VARIABLES (ANALYTICAL)

Description	Result	Implications
<u>RESEARCH QUESTION #1</u> Which one of the candidate distinguished segmentation bases, identified in phase I, is best predicted once we have the knowledge of a set of relevant independent variables?	BEHAVE (73.84%)	When the other segmentation variables are used as descriptors, energy related behavior is best predicted by these variables.
<u>RESEARCH QUESTION #2</u> How much of the uncertainty in the distinguished variable is reduced when it is dichotomized into a class (or group of classes) versus all of the other classes?	73.84% (Classes 1 & 2)	The classification of the energy related behavior to two classes of "action taken/action not taken," 73.84% of uncertainty regarding this variable is removed.
<u>RESEARCH QUESTION #3</u> How can one prioritize the effect of the independent variables on the distinguished set?	1:AWARE 2:KWHUSE 3:ENRATT 4:PERCEPT 5:GENATT	In reducing the number of descriptor variables to be used in subsequent analyses, use the given order (AWARE=Most-GENATT=least-important).
<u>RESEARCH QUESTION #4</u> What is the dependency structure in each class of the distinguished segmentation base?	KBA/EBK KEPGA/ABK	When four segmentation variables are used, indirect/direct effects in the simplest structure are as given. When all six segmentation variables are used, indirect/direct effects in the simplest structure are as given.

TABLE XXXIX

SIGNIFICANCE OF THE RESULTS OF SEGMENT
CONGRUENCE ANALYSIS TO THE PNWRES STUDY

PHASE II: EXOGENOUS DESCRIPTOR VARIABLES

Description	Result	Implications
<u>RESEARCH QUESTION #1</u> Which one of the candidate distinguished segmentation bases, identified in phase I, is best predicted once we have the knowledge of a set of relevant independent variables?	KWHUSE (13.18%)	When the exogenous descriptor variables are used, electricity consumption is the best predicted segmentation variable.
<u>RESEARCH QUESTION #2</u> How much of the uncertainty in the distinguished variable is reduced when it is dichotomized into a class (or group of classes) versus all of the other classes?	16.56% (4&5 vs. others) 9.13% (1&2 vs. others) 6.50% (3 vs. others)	The classification of the electricity consumption variables to dichotomous variables (light, medium, or heavy users vs others) shows the light user segment is best predicted by the exogenous variables.
<u>RESEARCH QUESTION #3</u> How can one prioritize the effect of the independent variables on the distinguished set?	1:DEMOG 2:TYPDWEL 3:CLIMGEO 4:INSUL 5:RENTOWN	In reducing the number of descriptor variables to be used in subsequent analyses, use the given order (DEMOG=Most-order (DEMOG=Most-RENTOWN=least-important).
<u>RESEARCH QUESTION #4</u> What is the dependency structure in each class of the distinguished segmentation base?	KT/TD/DC KCTDI/ KCRDI/ KCTRD	When only four variables are used, indirect/direct effects in the simplest structure are as given. When all six variables are used, indirect/direct effects in the simplest structure are as given.

These results entail several implications for managers at BPA or other utility companies. These implications are exemplified under the context of identifying different segments on the basis of electricity consumption as follows:

- o Electricity consumption is significantly interrelated with attitudes and energy related perceptions, awareness, and behavior.
- o Electricity consumption (in general) is somewhat better predicted by exogenous descriptor variables (13.18%), (i.e., demographics of the residents, type of dwelling unit, geoclimatic environment, level of insulation of the residence, and rental/ownership status of the residence) than by the other segmentation variables (10.99%).
- o The high electricity user is better predicted by the exogenous variables (16.56%) as compared with the medium and light users (6.50% and 9.13%, respectively).
- o Demographic characteristics of the residents, type of the residential unit, and the geoclimatic variables are the highest contributors to the knowledge of electricity consumption among the exogenous variables considered.

These, and similar interpretations of the segment congruence analysis will enhance the marketing manager's understanding of the issues surrounding segmentation studies. The ultimate decision, of which segment(s) to select for targeting purposes, is, of course up to the marketing manager.

CONTRIBUTIONS OF THE STUDY

The present study resulted in contributions in several areas including (a) segment congruence analysis, (b) market segmentation (c) log-linear modeling, and (d) information theoretic modeling. These contributions are examined in this section.

Contribution to Segment Congruence Analysis

Marketing practitioners should recognize that in many markets it is necessary to refine their definition of market segments in order to be able to serve more efficiently and effectively. To do so, they must recognize that segmentation based on a single, a priori selected variable will often result in vague and inaccurately defined (set of) segment(s). On the other hand, in many cases, segmentation variables are a set of discrete (or discretized) variables for which data has been collected on dichotomous or polytomous scales.

The above two characteristics of modern market segmentation (i.e., multiple variables and categorical data) bring about the need for techniques for analyzing such data (i.e., discrete multivariate analysis techniques). The study of the interrelations between a set of segmentation variables has been termed segment congruence analysis. Contemporary marketing researchers and practitioners have

advocated a variety of approaches to such analyses.

However, they have stopped after what can be considered an exploratory study of multiple segmentation variables (Green and Carmone 1977, Van Auken and Lonial 1984).

The present study extends segment congruence analysis to include several other important aspects of discrete multivariate analysis. This extension includes:

- o a number of alternative methods of identifying candidate segmentation variables,
- o objective identification of a distinguished segmentation variables (using endogenous or exogenous descriptor variables),
- o prioritization of the descriptor variables with respect to their contribution to the knowledge about the segmentation variable, and
- o structural modeling of the system of segmentation and descriptor variables with the purpose of understanding the nature and magnitude of interrelations among them.

These additional questions have enhanced segment congruence analysis from a method of exploratory analysis of systems of multiple segmentation variables to a descriptive study of such systems. Particularly, the addition of structural modeling techniques addresses the nature of interrelations among these variables.

Furthermore, the introduction of exogenous descriptor variables has enabled segment congruence analysis to be applied to an actual market segmentation study. In this study, the structure of the market is studied, not only to identify a distinguished segmentation variable, but, to

describe the relation of this variable to a set of background variables.

Contribution to Market Segmentation

Contributions of this research to the general area of market segmentation is threefold. First, as mentioned before, by introducing a set of exogenous descriptor variables, segment congruence analysis can be viewed as an integral part of market segmentation. Any segmentation study, a priori, a posteriori, or hybrid, can use phase II of this study.

Second, a posteriori and hybrid market segmentation studies can take advantage of the methods proposed for identifying candidates for the distinguished segmentation variable. Several researchers including Green (1977), Wind (1978), and others have expressed concern about a posteriori segmentation studies with respect to the meaningfulness of their results. A segmentation variable identified by clustering based techniques may produce highly significant results analytically, but completely anomalous results practically.

The present study, can not, of course, claim the prevention of such results. However, it can help to identify a more meaningful practical (as well as analytical) segmentation variable by proposing a number of them as the

candidates to be evaluated for the distinguished segmentation variable.

Third, by introducing the information theoretic approach to segment congruence analysis, this study has provided a powerful tool for analysis of discrete multivariate systems, which is extremely useful for modern market segmentation analysis. Information theory has already been used to develop brand-choice models by Herniter (1973 and 1974) and Bass (1976) among others. This study introduce this approach to market segmentation studies.

Contribution to Log-Linear Modeling

In its more than four decades of history, the development of information theory has enabled the formulation of many interesting managerial concerns in the form of easy-to-understand expressions and procedures. Log-linear models in their current form are relatively new, therefore, many models, with managerial implications, have not been developed explicitly, in spite of these techniques ability to address them. Furthermore, some of the commonly used models have been misinterpreted with respect to their meaning and the hypotheses tested.

Two misinterpretations of the log-linear models, which were encountered in this study are discussed next. This discussion is followed by two log-linear based procedures to address the fifth research questions in both phases.

The second research question in the first phase (i.e., contribution of each segmentation variable to the overall mutual association) has, traditionally, been assessed using the one-variable independence model (e.g., $\{A\}\{BCD\dots\}$). It was shown, in this study, that the correct model for this research question is the mutual independence model minus the two-variable independence models (i.e., $\{ABCD\dots\}-\{AB\}\{AC\}\dots$). This was done by comparing the two models with the information theoretic expression used for this question (i.e., $T_A(BCD\dots)=T(A:B:C:\dots)-T(A:B)-T(A:C)-\dots$). This was obviously a case of misinterpreting the log-linear models.

It was then shown that the one-variable independence model addresses the third research question in the first phase (i.e., the contribution of each variable to the joint variability in the system of segmentation variables). This, again was initially discovered by the fact the information theoretic model --e.g., $S(ABC\dots)-S_A(BC\dots)$, generated identical statistics to the one-variable independence model.

The second misinterpretation relates to the third research question in the second phase (i.e, prioritize the effect of the descriptor variables on the distinguished variable). Green and Carmone (1977), briefly, claim that the coefficients of the logit model can be used to prioritize the effect of the independent variables on the dependent variable variable. This is true only in the case

of all dichotomous independent variables. In this case each variable will have only one coefficient. (By analogy to the more commonly known dummy variable regression models, each dichotomous independent variable will have one dummy variable in the model).

However, in the case of polytomous independent variables, the model will produce several coefficients for each variable, which are absent or present, depending on which classes of the dependent variable are being considered. This problem was resolved by developing log-linear models (equivalent to Abrahamse and van Bueren, 1980) to assess the contributions of the descriptor variables to the segmentation variable.

Two log-linear procedures were developed to address two other research questions. The first procedure, which was used in addressing the fifth research question in the first phase, involved developing successive bivariate log-linear models. These models were, then, summarized in a two dimensional table (this table was, subsequently used to identify the variable with the highest number of dyadic interactions with other variables).

The second procedure was developed to construct equivalent models to Conant's (1981 and 1982) dependency analysis. The contributions of the other variables to each variable were assessed and a prioritization scheme similar to that of Abrahamse and van Bueren's (1980) was developed

for each variable. Then, these prioritizations were used to construct the approximate structure for the multivariate system. This approach, of course, produced the same exact structural model for the data.

Contribution to Information Theoretic Modeling

Information theoretic models have been used in economics, psychology, sociology, business administration, and other social sciences. However, most of these applications, particularly those to business administration (e.g., accounting (Lev 1968, 1970), finance (Philipatos and Gressis 1975), management (Horowitz and Horowitz 1976), and marketing (e.g., Herniter 1973 and 1974, Bass 1976)), have largely been simple applications of constraint analysis. This study applies models as simple as entropy and transmission to more complicated models, such as dependency analysis.

This study represents the first application of information to segment congruence analysis. It also represents the first in-depth application of this technique to any segmentation problem. Furthermore, this research project is among the very few undertakings in marketing which utilize information theoretic models.

It is through such applications and comparisons that a researcher can understand and appreciate the efficacy of an approach. Information theory seems to demonstrate

considerable capability in addressing the research questions addressed throughout this study.

LIMITATIONS

As any other research project, this study has certain limitations. These limitations, however, should not provide a hindrance in realizing the benefits provided by the proposed methodology.

This study is methodological, and therefore, emphasizes the techniques used, rather than the relevance of the application and the particular data analyzed (though, the analysis produced plausible results). The strongest motives for selecting the PNWRES database were the relevance of the data to the segment congruence issue, availability of the data, the large sample size, and the large number of variables measured.

Since the PNWRES data were not collected specifically for market segmentation purposes, some limitation arise. For instance, the survey was far more oriented toward the residences and their characteristics than toward the consumers. Although the residence descriptors are thought to be important precursors of individuals energy related behavior, a more detailed description of consumer characteristics would have been useful.

All variables used in an application oriented research project must be examined for reliability and scrutinized for validity. Validity and reliability issues were not addressed, as the purpose of this study was not to assess issues in the energy industry per se, but to demonstrate the relevance of the research questions and the methodology through a readily available data set. On the other hand, since many methodological studies use data which are (a) hypothetical (i.e., contrived) and/or (b) contain small number of observations, the 1983 PNWRES data had the added advantage of being an actual survey with 4703 cases. The results of this study, therefore, are far more indicative of the efficacy of the methods used.

Finally, the only means of the initial variable selection for the study was expert opinion. Initially a large number of variables were selected (a priori) as potential segmentation and descriptor variables. These variables were presented to the executives at BPA and PP&L. Their recommendations were accommodated and the final list of variables was compiled.

Again, it is emphasized that the analysis provided an excellent demonstration for the efficacy of the techniques used. Furthermore, the results of the application did not seem to produce counter-intuitive conclusions despite the mentioned limitations.

DIRECTIONS FOR FURTHER RESEARCH

Several directions for further research can readily be identified. These directions can be categorized under (1) segment congruence analysis, (2) log-linear modeling, (3) information theoretic modeling, and (4) other.

Segment congruence analysis is a relatively new area in market segmentation. The present study attempted to develop an exploratory and, to an extent, a descriptive approach to segment congruence analysis. The natural continuation of this stream of research would involve development of models for predictive purposes. Both information theory and log-linear models offer rich prospects in this area. Also, an application oriented research project, with primary data collection, is needed in order to further establish segment congruence analysis as a practical and valuable tool.

Log-linear modeling is relatively young, particularly, in the area of structural modeling. Both researchers and practitioners would benefit from further development in this area. Continuing with comparisons between better known models, such as those offered by information theory, would provide an excellent stream of research. Information theory contains models that deal with phenomena in the context of Bayesian statistics. It would probably be fruitful to develop the log-linear approach to such models.

Information theoretic models from simple models of marginal and conditional entropies, transmissions, interactions, and systematic entropies to the complex models for spectral analysis, dependency analysis and reconstructability analysis, offer a rich family of techniques applicable to many business problems. The bridge between business application and information theory must be strengthened and maintained. This would set the stage for synergistic developments in business problem solving and statistical modeling.

Information theoretic structural modeling was proposed by Ashby (1964) and, more recently, further developed by Klir and his colleagues, and several other researchers including Broekstra, Krippendorff, and Conant. The present study employed Conant's (1981, 1982) dependency analysis. The other methods for structural modeling, proposed by these authors, can be incorporated in further development of segment congruence analysis.

Finally, it is highly desirable to develop an "expert system" for segment congruence analysis through the two methodologies discussed in this study. In this era of information explosion and machine intelligence, developing such a system would help reduce the immense complexity surrounding managers and decision makers, as well as taking advantage of the emerging power of machines.

BIBLIOGRAPHY

- Abrahamse, A.P.G., and W.M. Lammerts van Bueren (1980)
"Nominal Variables: Dependence and Predictive Power,"
Report 8017/S, Economic Institute, Erasmus University
Rotterdam.
- Alpert, M.I. (1971) "A Canonical Analysis of Personality and
the Determinants of Automobile Choice," Proceedings,
American Marketing Association, 312-316.
- Andrews, F.M., and R.C. Messenger (1973) Multivariate
Nominal Scale Analysis, Ann Arbor, Michigan: Survey
Research Center, University of Michigan.
- Andrews, Frank M., Morgan James N., and Sonquist John A.,
(1969) Multiple Classification Analysis, Survey Research
Center, Institute for Social Research, The University of
Michigan, Ann Arbor, Michigan.
- Arabie, P. (1977) "Clustering Representations of Group
Overlap," Journal of Mathematical Sociology, 5, 113-28.
- Arabie, P., J.D. Carroll, W. DeSarbo, and J. Wind (1981)
"Overlapping Clustering: A New Method for Product
Positioning," Journal of Marketing Research, XVIII,
(August) 310-17.
- Armstrong, J.S., and J.G. Anders (1970) "Exploratory
Analysis of Marketing Data: Trees vs. Regression,"
Journal of Marketing Research, VII, (November) 487-492.
- Ashby, W.R. (1964) "Constraint Analysis of Many-valued
Relations," General Systems, 9, 99-105.
- Assael, H., (1970) "Segmenting Markets by Group Purchasing
Behavior: An Application of AID Technique," Journal of
Marketing Research, VII, (May) 153-158.
- Attneave, F. (1959) Applications of Information Theory to
Psychology: A Summary of Basic Concepts, Methods, and
Results, New York: Holt, Rinehart and Winston.

- Bartlett, M.S. (1955) An Introduction to Stochastic Processes, Cambridge Univ. Press.
- Bass, F.M. (1976) "The Theory of Stochastic Preference and Brand Switching," Journal of Marketing Research, XI (February) 1-20.
- Batsell, R.R. (1980) "Consumer Resource Allocation Models at the Individual Level," Journal of Consumer Research, 7 (June) 78-87.
- Batsell, R.R., and L.M. Lodish (1981) "A model and Measurement Methodology for Predicting Individual Consumer Choice," Journal of Marketing Research, 18 (February) 1-12.
- Bernhardt, I. and K.D. MacKenzie (1968) "Measuring Seller Unconcentration, Segmentation and Product differentiation," Western Economic Journal, 6 (December) 395-403.
- Bernhardt, K.L., and T.C. Kinnear (1976) "Categorical Regression in Marketing," Journal of Business Research, 4, (November) 297-312.
- Best, R.J. (1975) The Relationship Between Joint-Space Distance and Stochastic Choice: An Empirical Investigation, Ph. D. Dissertation, University of Oregon, Eugene, Oregon.
- Bishop, Y.M.M. (1969) "Full Contingency Tables, Logits, and Split Contingency Tables," Biometrics, (June) 383-399.
- Bishop, Y.M.M., S.E. Fienberg, and P.W. Holland (1975) Discrete Multivariate Analysis: Theory and Practice, Cambridge, Mass.: The MIT Press.
- Blalock, Hubert M., (1972) Social Statistics, McGraw-Hill Book Company.
- Blattberg, R., and S. Sen (1976) "Market Segmentation and Stochastic Brand Choice Models," Journal of Marketing Research, XIII, (February) 34-45.
- Brillouin, L. (1956) Science and Information Theory, New York: Academic Press.
- Broekstra, G. (1978) "On the Representation and Identification of Structure Systems," International Journal of Systems Science, 9, 1271-1293.

- Broekstra, G. (1981) "C-Analysis of C-Structures: Representation and Evaluation of Reconstruction Hypotheses by Information Measures," International Journal of Systems, 7, 33-61.
- Broekstra, G. (1982) "Constraint Analysis of Discrete Multivariate Systems," Proceedings 26th Annual Meeting of the Society for General Systems Research, Washington D.C., 259-265.
- Cavallo, R.E. and G.J. Klir (1979) "Reconstructability Analysis of Multi-Dimensional Relations: A Theoretical Basis for Computer-Aided Determination of Acceptable Systems Models," International Journal of General Systems, 5, 143-171.
- Cavallo, R.E. and G.J. Klir (1981) "Reconstructability Analysis: Evaluation of Reconstruction Hypotheses," International Journal of General Systems, 7, 7-32.
- Chapman, R.G. and R. Staelin (1982) "Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model," Journal of Marketing Research, 19 (August) 288-301.
- Cherry, E.C. (1952) "The Communication of Information", American Scientist, 40, 640-664.
- Clogg, C.C. and J.M. Munch (1984) "Using Simultaneous Latent Structure Models to Analyze Group Differences: Exploratory Analysis of Buying Style Items," Journal of Business Research, 12, 319-336.
- Clogg, C.C., and L.A. Goodman, (1985) "Simultaneous Latent Structure Analysis in Several Groups," Sociological Methodology, 59, 124-38.
- Coleman, J.S. (1964) Introduction to Mathematical Sociology, New York: Free Press.
- Conant, R.C. (1974) "Information Flows in Hierarchical Systems," International Journal of General Systems, 1, 9-18.
- Conant, R.C. (1981) "Detection and Analysis of Dependency Structures," International Journal of General Systems, 7, 81-91.
- Conant, R.C. (1982) "Approximating General Structures with Dependency Analysis," Proceedings 26rd Annual Meeting of the Society for General Systems Research.

- Cramer, H. (1946) Mathematical Methods of Statistics, Princeton Univ. Press.
- Currim, I.S., (1981) "Using Segmentation Approaches for Better Prediction and Understanding from Consumer Mode Choice Models," Journal of Marketing Research, 18 (August) 301-9.
- Currim, I.S., (1982) "Predictive Testing of Consumer Choice Models That Are Not Subject to Independence of Irrelevant Alternatives," Journal of Marketing Research, 19 (May) 208-22.
- Darroch, J.N. (1962) "Interactions in Multifactor Contingency Tables," Journal of the Royal Statistical Society, 24, 251-263.
- Davis, J.A., (1975) "Analyzing Contingency Tables with Linear Flow Graphs: D Systems," In Sociological Methodology 1976, edited by D.R. Heise, 111-45. San Francisco: Jossey-Bass.
- Dempster, A.P., N.M. Laird, and D.B. Rubin (1977) "Maximum Likelihood From Incomplete Data Via the EM algorithm," Journal of the Royal Statistical Society, Series B, 39, 1-38.
- DeSarbo, W.S., and D.K. Hildebrand (1980) "A Marketer's Guide to Log-Linear Models for Qualitative Data Analysis," Journal of Marketing, (Summer) 44, 40-51.
- Fisher, R.A. (1925) "Theory of Statistical Estimation," Proceedings of Cambridge Philosophical Society, 22, 700-725.
- Flath, D. and E.W. Leonard (1979) "A Comparison of Two Logit Models in the Analysis of Qualitative Marketing Data," Journal of Marketing Research, 16 (November) 533-38.
- Frank, R.E., W.F. Massy, and Y. Wind, (1972) Market Segmentation, Prentice-Hall, Inc., Englewood Cliffs, N.J.
- Friedlander, D. (1961) "A Technique for Estimating a Contingency Table Given the Marginal Totals and Some Supplementary Data," Journal of the Royal Statistical Society, 124, 412-420.
- Fritzsche, D.J. (1981) "An Analysis of Energy Consumption Patterns by Stage of Family Life Cycle," Journal of Marketing Research, XVIII (May) 227-232.

- Garner, W.R. (1962) Uncertainty and Structure as Psychological Concepts, New York: John Wiley and Sons, Inc.
- Garner, W.R. and W.J. McGill (1956) "The Relation Between Information and Variance Analysis, Psychometrika, 21 (September) 219-228.
- Gensch, D.H. and W.W. Recker (1979) "The Multinomial, Multiattribute Logit Choice Model," Journal of Marketing Research, 16 (February) 124-32.
- Goodman, L.A. (1963) "On Methods for Comparing Contingency Tables," Journal of the Royal Statistical Society, Series A, 126, 94-108.
- Goodman, L.A. (1964a) "Simultaneous Confidence Limits for Cross-Product Ratios in Contingency Tables," Journal of the Royal Statistical Society, Series B, 26, 86-102.
- Goodman, L.A. (1964b) "Simple Methods for Analyzing Three-Factor Interaction in Contingency Tables," Journal of the American Statistical Association, 59, 319-52.
- Goodman, L.A. (1964c) "Interactions in Multidimensional Contingency Tables," Annals of Mathematical Statistics, 35, 632-46.
- Goodman, L.A. (1964d) Simultaneous Confidence Intervals for Multinomial Populations," Annals of Mathematical Statistics, 35, 716-25.
- Goodman, L.A. (1965a) "On Simultaneous Confidence Intervals for Multinomial Proportions," Technometrics, 7, 247-54.
- Goodman, L.A. (1965b) "On the Statistical Analysis of Mobility Data," American Journal of Sociology, 70, 564-85.
- Goodman, L.A. (1965c) "On the Multivariate Analysis of Three Dichotomous Variables," American Journal of Sociology, 71, 290-301.
- Goodman, L.A. (1969a) "How to Ransack Social Mobility Tables and Other Kinds of Cross-Classification Tables," American Journal of Sociology, 75, 1-40.
- Goodman, L.A. (1969b) "On Partitioning X^2 and Detecting Partial Associations in Three-Way Contingency Tables," Journal of the Royal Statistical Society, Series B, 31, 486-98.

- Goodman, L.A. (1970) The Multivariate Analysis of Qualitative Data," Journal of the American Statistical Association, 65, 226-56.
- Goodman, L.A. (1971a) "The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications," Technometrics, 13, 33-61.
- Goodman, L.A. (1971b) "Partitioning of Chi Square, Analysis of Marginal Contingency Tables, and Estimation of Expected Frequencies in Multidimensional Contingency Tables," Journal of the American Statistical Association, 66, 339-44.
- Goodman, L.A. (1968) "The Analysis of Cross-Classified Data: Independence, Quasi-Independence, and Interactions in Contingency Tables with or without Missing Entries," Journal of the American Statistical Association, 63, 1091-1131.
- Goodman, L.A. (1972) "A General Model for the Analysis of Surveys," American Journal of Sociology, 77, 1035-1087.
- Goodman, L.A. (1973) "The Analysis of Multidimensional Contingency Tables When Some Variables are Posterior To Others: A Modified Path Analysis Approach," Biometrika, 59, 579-96.
- Goodman, L.A., (1974) "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models," Biometrika, 61, 215-31.
- Goodman, L.A. (1978) Analyzing Qualitative/Categorical Data, Edited by J. Magidson, Abt Books, Cambridge, Massachussettes.
- Green, P.E. (1977) "A New Approach to Market Segmentation," Business Horizons, (February) 61-73.
- Green, P.E. (1978) "An AID/Logit Procedure for Analyzing Large Multiway Contingency Tables," Journal of Marketing Research, XV, (February) 132-36.
- Green, P.E. and F.J. Carmone (1977) "Segment Congruence Analysis: A Method for Analyzing Association Among Alternative Bases for Market Segmentation," Journal of Consumer Research, 3 (March) 217-222.
- Green, P.E., and V.R. Rao (1972) Applied Multidimensional Scaling: A Comparison of Approaches and Algorithms, New York: Holt, Rinehart and Winston.

- Green, P.E., F.J. Carmone, and D.P. Wachspress (1976) "Consumer Segmentation Via Latent Class Analysis," Journal of Consumer Research, 3 (December) 170-174.
- Green, P.E., F.J. Carmone, and D.P. Wachspress (1977) "On the Analysis of Qualitative Data in Marketing Research," Journal of Marketing Research, XIV (February) 52-59.
- Grizzle, J.E., C.F. Stamper, and G.G. Koch, (1969) Analysis of Categorical Data by Linear Models," Biometrics, 25, 489-504.
- Haberman, S.J. (1972) "Log-linear Fit for Contingency Tables," Applied Statistics, 21, 218-25.
- Haberman, S.J. (1982) "Analysis of Dispersion of Multinomial Response," Journal of the American Statistical Association, 77, No. 379 (September) 568-580.
- Harmon, R.R., J. Hosseini, and M. Zwick (1986) "Application of Information Theory to the Analysis of Market Segments when Multiple Segmentation Variables Are Used," Proceedings, American Marketing Association, Educators Meeting, Annual Conference, (August).
- Haberman, S.J. (1979) Analysis of Qualitative Data, 1&2, New York: Academic Press.
- Herniter, J.D. (1973) "An Entropy Model of Brand Purchase Behavior," Journal of Marketing Research, X (November) 361-375.
- Herniter, J.D. (1974) "A Comparison of the Entropy Model and the Hendry Model," Journal of Marketing Research, XI (February) 21-29.
- Horowitz, A.R. and I. Horowitz (1976) "Education: The Real and Illusory Virtues of Entropy-Based Measures for Business and Economic Analysis," Decision Sciences, 7, 121-136.
- Hosseini, J.C., R.R. Harmon, (1986) "Analyzing Overlapping Market Segments: An Information Theoretic Approach," Proceedings, Western American Institute for Decision Sciences, Winter Conference, (March).
- Hosseini, J.C., R.R. Harmon, and M. Zwick (1986) "Segment Congruence Analysis Via Information Theory," Proceedings, Society for General Systems Research, Annual Conference, (May)

- Hauser, J.R., and G.L. Urban (1977) "A Normative Methodology for Modeling Consumer Response to Innovation," Operations Research, 25 (July-August) 579-619.
- Jones, J.M. and F.S. Zufryden (1980) "Adding Explanatory Variables to a Consumer Purchase Behavior Model: An Exploratory Study," Journal of Marketing Research, 17 (August) 323-34.
- Klir, G.J. (1976) "Identification of Generative Structures in Empirical Data," International Journal of General Systems, 3, 89-104.
- Klir, G.J. (1986) "Third W. Ross Ashby Memorial Lecture 1986. Reconstructability Analysis: An Offspring of Ashby's Constraint Analysis," Systems Research, 3, No. 4, 267-271.
- Knoke, D., and P.J. Burke (1980) "Log-Linear Models," Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-020, Beverly Hills and London: Sage Publications.
- Krippendorff, K. (1979) "On the Identification of Structures in Multivariate Data by the Spectral Analysis of Relations," Proceedings 23rd Annual Meeting of the Society for General Systems Research, Austin, Texas.
- Krippendorff, K. (1981) "An Algorithm for Identifying Structural Models of Multi-Variate Data," International Journal of General Systems, 7, 63-79.
- Krippendorff, K. (1986) "Information Theory: Structural Models for Qualitative Data," Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-062, Beverly Hills and London: Sage Publications.
- Ku, H.H. and S. Kullback (1968) "Interaction in Multidimensional Contingency Tables: An Information Theoretic Approach," Journal of Research of the National Bureau of Standards, 72B (July-September) 159-99.
- Kullback, S. (1959) Information Theory and Statistics, New York: John Wiley and Sons, Inc.

- LaGarce, R. (1974) "An Analysis of Second-Order Interaction in Multidimensional Contingency Tables," Journal of Marketing Research, 11 (August) 343-45.
- Lev, B. (1968) "The Aggregation Problem in Financial Statements: An Information Approach," Journal of Accounting Research, 6, (Autumn) 247-61.
- Lev, B. (1970) "The Informational Approach to Aggregation in Financial Statements: Extensions," Journal of Accounting Research, 8, (Spring) 78-94.
- Lazarsfeld, P.F. and N.W. Henry (1942) Latent Structure Analysis, Boston, Houghton Mifflin Company.
- Lewis, B.N., (1962) "On the Analysis of Interaction in Multi-Dimensional Contingency Tables," Journal of Royal Statistical Society, Series A, 125, 88-117.
- Lilien, G.L. and P. Kotler (1983) Marketing Decision Making: A Model-Building Approach, New York: Harper and Row.
- Malhotra, N.K. (1982a) "Information Load and Consumer Decision Making," Journal of Consumer Research, 8 (March) 419-30.
- Malhotra, N.K. (1982b) "Structural Reliability and Stability of Non-Metric Conjoint Analysis," Journal of Marketing Research, 19 (May) 199-207.
- Malhotra, N.K., A.K. Jain, and S. Lagakos (1982) "The Information Load Controversy: An Alternative Viewpoint," Journal of Marketing, 46 (Spring) 27-37.
- Martin, C.R., and R.L. Wright (1974) "Profit Oriented Data Analysis for Market Segmentation: An Alternative to AID," Journal of Marketing Research, XI, (August) 399-412.
- McGill, W.J. (1954) "Multivariate Information Transmission," Psychometrika, 19, 97-116.
- Miller, G.A. (1953) "What is Information Measurement?," American Psychologist, 8, 3-11

- Miller, J., and M. Erickson, (1974) "On Dummy Variable Regression Analysis: A Description and Illustration of the Method," Sociological Methods and Research, 2, 409-30.
- Mitra, S.K. (1955) "Contributions to the Statistical Analysis of Categorical Data," N.C. Institute of Statistics, Mimeo Series No. 142 (December).
- Morgan James N. and Messenger Robert C., (1973) THAID: A Sequential Program for analysis of Nominal Scale Dependent Variables, Survey Research Center, Institute for Social Research, The University of Michigan, Ann Arbor, Michigan.
- Myers, J.G. and F.M. Nicosia (1967) "New Empirical Directions in Market Segmentation: Latent Structure Models", AMA Winter Conference Proceedings, Chicago, Ill.: American Marketing Association.
- Newman, J.W., and R. Staelin (1971) "Multivariate Analysis of Differences in Buyer Decision Time," Journal of Marketing Research, VIII, (May) 192-8.
- Pearson, K. (1904) "Mathematical Contributions to the Theory of Evolution, XIII, on the Theory of Contingency and Its Relation to Association and Normal Correlation," Drap. Co. Mem. Biom. Ser., 1.
- Perreault, Jr., W.D. and H.C. Barksdale, Jr. (1980) "A Model-Free Approach for Analysis of Complex Contingency Data in Survey Research," Journal of Marketing Research, XVII (November) 503-15.
- Perreault, Jr., W.D., D.N. Behrman, and G.M. Armstrong (1979) "Alternative Approaches for Interpretation of Multiple Discriminant Analysis in Marketing Research," Journal of Business Research, 7, 151-173.
- Peters, W.H. (1970) "Using MCA to Segment New Car Markets," Journal of Marketing Research, VII, (August) 360-3.

- Philippatos, G.C., and N. Gressis (1975) "Conditions of Equivalence Among E-V, SSD, and E-H Portfolio Selection Criteria: The Case for Uniform, Normal and Lognormal Distributions," Management Science, 21, (February) 617-25.
- Punj, G.N., and R. Staelin (1978) "The Choice Process for Graduate Business Schools," Journal of Marketing Research, 15 (November) 588-98.
- Rao, C.R. (1952) Advanced Statistical Methods in Biometric Research, New York: John Wiley and Sons.
- Rao, C.R. (1965) "Criteria of Estimation in Large Samples," 345-362 in Contributions to Statistics, New York: Pergamon Press.
- Rao, V.R. and F.W. Winter (1978) "An Application of the Multivariate Probit Model to Market Segmentation and Product Design," Journal of Marketing Research, XV (August) 361-368.
- Rao, V.R. and H.S. Solgaard (1977) "A Comparison of Qualitative Response Models of Consumer Choice," AMA Educator's Proceedings, Chicago, Ill.: American Marketing Association, 135-140.
- Reynolds, H.T. (1977) "Analysis of Nominal Data," Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-007, Beverly Hills and London: Sage Publications.
- Roy, S.N. and M.A. Kastenbaum (1955) "A Generalization of Analysis of Variance and Multivariate Analysis to Data Based on Frequencies in Qualitative Categories or Class Intervals," N.C. Institute of Statistics, Mimeo Series No. 131 (June 1).
- Russell, S.W. and M.J. Rayan (1979) "Analyzing Cross-Classification Data: An Improved Method for Predicting Events," Journal of Marketing Research, XVI, (November) 539-544.
- Shannon, C.E. and W. Weaver (1949) The Mathematical Theory of Communication, University of Illinois Press, Urbana, Chicago, London.

- Silk, A.J., and G.L. Urban (1978) "Pre-Test-Market Evaluation of New Packaged Goods: A Model and Measurement Methodology," Journal of Marketing Research, 15 (May) 171-91.
- Simpson, C.H., (1951) "The Interpretation of Interaction in Contingency Tables," Journal of Royal Statistical Society, Series B, 13, 238-41.
- Smith, W.R. (1956) "Product Differentiation and Market Segmentation as Alternative Marketing Strategies," Journal of Marketing, (July) 21, 3-8.
- SPSS^X Inc., (1986) SPSS^X User's Guide, 2nd Ed., McGraw-Hill Book Company.
- Stouffer, S.A., E.A. Schuman, L.C. Devinney, S.A. Star, and R.M. Williams, Jr. (1949) The American Soldier: Adjustments During Army Life. Studies in Social Psychology in World War II, 1, Princeton N.J.: Princeton University Press.
- Stumpers, F.L.H.M. (1953) "A Bibliography of Information Theory; Communication Theory-Cybernetics," R.L.E., M.I.T. (February 2) IRE Trans., PGIT-2 (November).
- Theil, H. (1967) Economics and Information Theory, Chicago: Rand McNally and Company.
- Theil, H. (1970) "On the Estimation of Relationships Involving Qualitative Variables," American Journal of Sociology, 76, 103-154
- U.S. Department of Energy, Bonneville Power Administration, (1983) PNWRES User's Guide, (Spring).
- Van Auken, S., and S.C. Lonial (1984) "Assessing Mutual Association Between Alternative Market Segmentation Bases," Journal of Advertising, 13, No. 1, 11-16.
- von Bertalanffy, L. (1968) General System Theory: Foundations, Development, Applications, New York, George Braziller.
- Watanabe, S. (1969) Knowing and Guessing, John Wiley and Sons, Inc.
- Wilks, S.S. (1935) "The Likelihood Test of Independence in Contingency Tables," Annals of Mathematical Statistics, 6, 190-196.

- Wilks, S.S. (1943) Mathematical Statistics, Princeton Univ. Press.
- Wilner, D.M., R.P. Walkley, and S.W. Cook (1955) Human Relations in Interracial Housing: A Survey of the Contact Hypothesis, Minneapolis: University of Minnesota Press.
- Wind, Y. (1978) "Issues and Advances in Segmentation Research," Journal of Marketing Research, XV, (August) : 317-337.
- Yule, G.U. and M.G. Kendall (1937) An Introduction to the Theory of Statistics, 11th ed., London: Charles Griffin.
- Zeisel, H. (1968) Say it with Figures, Fifth Edition, Revised, New York: Harper & Row.
- Zwick, M. (1979) "Information Theoretic Definition of "Leading Part"," Working Paper, System Science Program, Portland State University.

APPENDIX I

GENERAL LOG-LINEAR MODELS

There are two types of log-linear models; the general log-linear model and the logit model. In the general log-linear model variables are treated similarly with respect to their interdependence (i.e., no distinction is made based on dependence/independence between variables). In the logit model, the objective is prediction of a dependent variable from a set of independent variables.

The main vehicle of the log-linear models is the odds-ratio. TABLE 1 represents a 2x2 contingency tables. The odds ratio for this table is calculated as

$(F_{11}/F_{21})/(F_{12}/F_{22})$, or equivalently, $(F_{11}F_{22}/F_{12}F_{21})$.

While an odds-ratio of one indicates no association, this measure takes on any positive value. An odds-ratio smaller than one indicates an inverse relationship between the odds of F_{11} to F_{21} vs. those of F_{12} to F_{22} (i.e., a negative correlation), while an odds-ratio greater than one would mean a positive correlation in the above sense.

Assessment of any cross-tabulation through the log-linear models requires the odds-ratios as the basic form of

variation which needs to be explained. The general log-linear model develops a number of "reconstruction hypotheses" (or models) for the frequencies in a contingency table. Then the model which has the simplest form (to be defined later) and reconstructs the cell frequencies to an acceptable precision (as determined by the chi-square test of significance) is selected as the model for the data. For instance, the "saturated model" (i.e., the model which includes all possible interactions) would always reconstruct the data perfectly. A two-variable version of this model is discussed next with the purpose of demonstrating the essence of the general log-linear model.

The Saturated Model for a 2x2 Contingency Table

Consider a 2x2 table with f_{ij} representing the actual frequencies (see TABLE XXXX (a)). The saturated model for this contingency table (i.e., the model which represents all effects, such as the row effects, the column effects, and all interaction effects) would propose a set of expected cell frequencies, F_{ij} 's, (see Table XXXX (b)) exactly the same as the actual cell frequencies, F_{ij} . The reconstruction hypothesis is then stated as:

$$(26) \quad F_{ij} = y \cdot y_i^A \cdot y_j^B \cdot y_{ij}^{AB}$$

where;

$$y = (F_{11} \cdot F_{12} \cdot F_{21} \cdot F_{22})^{1/4}$$

$$y_1^A = (y_2^A)^{-1} = (F_{11} \cdot F_{12})^{1/2} / y$$

$$y_1^B = (y_2^B)^{-1} = (F_{11} \cdot F_{21})^{1/2} / y$$

$$y_{11}^{AB} = y_{22}^{AB} = (y_{12}^{AB})^{-1} = (y_{21}^{AB})^{-1}$$

$$= (F_{11} \cdot F_{22} / F_{12} \cdot F_{21})^{1/4}$$

TABLE XXXX

ACTUAL AND EXPECTED CELL FREQUENCIES FOR A
2x2 CONTINGENCY TABLE

a) ACTUAL CELL FREQUENCIES:

	Y	
X	f_{11}	f_{12}
	f_{21}	f_{22}

b) EXPECTED CELL FREQUENCIES:

	Y	
X	F_{11}	F_{12}
	F_{21}	F_{22}

c) A NUMERICAL EXAMPLE:

	Y	
X	50	80
	20	100

A numerical example of this model, based on Table
XXXX.(c), is presented below:

$$y = 53.183,$$

$$y_1^A = 1.189, \quad y_2^A = 0.841,$$

$$\begin{aligned}
y_1^B &= 0.595, & y_2^B &= 1.681 \\
y_{11}^{AB} &= 1.330, & y_{12}^{AB} &= 0.752, \\
y_{21}^{AB} &= 0.752, & y_{22}^{AB} &= 1.330
\end{aligned}$$

Then:

$$\begin{aligned}
F_{11} &= y \cdot y_1^A \cdot y_1^B \cdot y_{11}^{AB} = 50 \\
F_{12} &= y \cdot y_1^A \cdot y_2^B \cdot y_{12}^{AB} = 80 \\
F_{21} &= y \cdot y_2^A \cdot y_1^B \cdot y_{21}^{AB} = 20 \\
F_{22} &= y \cdot y_2^A \cdot y_2^B \cdot y_{22}^{AB} = 100
\end{aligned}$$

It can be shown that y , y_1^A , y_1^B , and y_{11}^{AB} are all based on the odds-ratio. For example, the row-effects odds ratio is stated as:

(27)

$$\begin{aligned}
(F_{11}/F_{21})(F_{12}/F_{22}) &= \frac{y \cdot y_1^A \cdot y_1^B \cdot y_{11}^{AB}}{y \cdot y_2^A \cdot y_1^B \cdot y_{21}^{AB}} * \frac{y \cdot y_1^A \cdot y_2^B \cdot y_{12}^{AB}}{y \cdot y_2^A \cdot y_2^B \cdot y_{22}^{AB}} \\
&= (y_1^A/y_2^A)^2 = (y_1^A)^4 = (1/y_2^A)^4
\end{aligned}$$

therefore;

$$\begin{aligned}
y_1^A &= 1/y_2^A = (F_{11}F_{12}/F_{21}F_{22})^{1/4} \cdot (F_{11}F_{12}/F_{11}F_{11})^{1/4} \\
&= (F_{11}F_{12})^{1/2}/F_{11}F_{12}F_{21}F_{22} = (F_{11}F_{12})^{1/2}/y
\end{aligned}$$

Other Models

The saturated models assumes that all effects (i.e., column, row, effects and joint two-way, three-way, etc. interaction effects) are present in a model. It is,

however, conceivable to eliminate some of these effects by setting them equal to one (e.g., in the case of interactions, setting $y_{11}^{AB} = 1$ would follow that $y_{ij}^{AB} = 1$ for all i and j which means A and B are independent). Then, this model would be acceptable if the generated expected frequencies (i.e., F_{ij} 's) are statistically the same as the actual frequencies (i.e., using the chi-square test for independence). Thus for the two dimensional contingency table (TABLE 2) the following models can be hypothesized:

$$\begin{aligned}
 (28) \quad F_{ij} &= y * y_i^A * y_j^B * y_{ij}^{AB} \\
 F_{ij} &= y * y_i^A * y_j^B \\
 F_{ij} &= y * y_i^A \\
 F_{ij} &= y * y_j^B \\
 F_{ij} &= y
 \end{aligned}$$

If these models are converted to their log-linear form, their interpretation becomes easier. For instance, taking the natural logarithm of all terms in the saturated model, $F_{ij} = y * y_i^A * y_j^B * y_{ij}^{AB}$, will yield:

$$(29) \quad \log (F_{ij}) = \log y + \log y_i^A + \log y_j^B + \log y_{ij}^{AB}$$

which can be re-written as:

$$(30) \quad \log F_{ij} = u + u_i^A + u_j^B + u_{ij}^{AB}$$

The above expression can be given a regression-like interpretation with u as the intercept, u_i^A as the effect of

variable A alone, u_j^B as the effect of variable B alone, and u_{ij}^{AB} as the joint effect of variables A and B.

Hierarchical Log-Linear Models

A very common version of the general log-linear model known as the hierarchical structure assumes that each model having a specific interaction term (e.g., u_{ij}^{AB}) includes all of the lower level terms (e.g., u , u_i^A , u_j^B). In this type of structure, the notation can be simplified. For example, the saturated two-variable model can be represented by {AB}. Similarly a multivariate model such as:

$$\begin{aligned}
 (31) \quad \log F_{ij} = & u + u_i^A + u_j^B + u_k^C + u_l^D \\
 & + u_{ij}^{AB} + u_{ik}^{AC} + u_{il}^{AD} + u_{jk}^{BC} + u_{jl}^{BD} \\
 & + u_{kl}^{CD} + u_{ijk}^{ABC} + u_{ijl}^{ABD} + u_{jkl}^{BCD}
 \end{aligned}$$

can be represented by {ABC}{ABD}{BCD}. In addition to simplifying the representation of a certain model, the above notation designates the subtables that need to be analyzed instead of the original contingency table.

Generating the Expected Frequencies

For a 2x2 table such TABLE 2, the expected frequencies are easily calculated. However, for more complicated tables, the log-linear models fall into two categories: Those with explicit formulae for generating expected frequencies (the direct or closed formed) and those without

such formulae (the indirect or open formed). (Reynolds, 1977).

In order to detect whether a model is closed or open formed, Reynolds (1977) offers a five step procedure (and also both Broekstra, 1981, and Krippendorff, 1981, offer a similar procedure when encountering the same problem in the information theoretic context). Then, to generate the expected frequencies for the open formed models an iterative method is selected --e.g., the Newton-Raphson or the Deming-Stephen algorithms (Knoke and Burke, 1980). After the expected frequencies are calculated for the proposed models, they are tested for independence using chi-square. The model with the lowest chi-square is then selected as the best model, as it reconstructs the data most closely.

APPENDIX II

ENTROPY AND INFORMATION THEORETIC CONCEPTS

As is well known, the foundations of mathematical information theory were laid down by Shannon (Shannon and Weaver 1949) during his work on the optimization of communication systems. Since then, this concept has made its way into the literature of Economics (Theil 1967), Psychology (Attneave 1959, Miller 1953, Garner 1962), Marketing (Herniter 1973, 1974 Bass 1976), Accounting (Lev 1968, 1970), Finance (Philipatos and Gressis 1975), and other disciplines.

It has been well established that the amount of potential information obtained from observing any state x of some phenomenon X can be measured by $-\log P_x$; where P_x is the probability of X being in state x . In general, however, all states of a variable do not have equal prior probabilities. Therefore, the amount of potential information to be gained from state x has to be weighted by the probability of its occurrence. If successive states are chosen independently, with fixed probabilities P_x , then a

convenient measure of the average amount of information gained with the knowledge of each state is:

$$(32) \quad H(X) = - \sum_{x=1}^N P_x \log P_x.$$

Equation (32) is known as the amount of information or the entropy of the source, because of the formal similarity of (32) with the thermodynamic definition of entropy (Shannon and Weaver 1949). The choice of a logarithmic base for Equation (32) is arbitrary. Commonly, in information theory, logarithms to base two are used, and information units are then measured in BInary digiTs (or BITS). Equation (32) can be extended to include both the multivariate case and the case with continuous distribution.

In the multivariate case, (32) can be rewritten as:

$$(33) \quad H(A,B,\dots,X) = - \sum_{a \in A} \sum_{b \in B} \dots \sum_{x \in X} P_{ab\dots x} \log P_{ab\dots x};$$

if the variable is continuous, then (32) is:

$$(34) \quad H(f(X)) = - \int f(x) \log f(x) dx.$$

The entropy of a continuous variable behaves in much the same way as does the entropy of a discrete variable.

The entropy $H(X)$ can be interpreted as a measure of the variability (scatter, spread, uncertainty, behavioral

The entropy $H(X)$ can be interpreted as a measure of the variability (scatter, spread, uncertainty, behavioral freedom) of the variable. 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 (Herniter 1973). It is a non-negative quantity applicable to metric and non-metric 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 freedom of choice is obtained when $H(X)$ is calculated from a uniform distribution (i.e., when all the states are equally probable), in which case $H(X) = \log_2 N$; N = the number of possible states for X .

In a similar vein, a system with two variables, X and Y , can be conceived as having total (joint) behavioral freedom, indicated by $H(X,Y)$. Intuition suggests that this quantity, $H(X,Y)$, is maximum if variables X and Y behave independently. In fact, this maximum possible value; denoted as $H_{\max}(X,Y)$, can be shown to be equal to the sum of the individual freedoms for variables X and Y (Broekstra 1981). That is;

$$(35) \quad H_{\max}(X,Y) = H(X) + H(Y).$$

On the other hand, if the two variables are mutually associated, their joint behavioral freedom is reduced, as they impose behavioral constraints upon one another.

We will next discuss what is known as conditional entropy and "transmission." For a system consisting of two variables X and Y , say $A = \{X, Y\}$, the variability of A , which is not accounted for in some subset of A , e.g., $\{Y\}$, is measured by the conditional entropy defined as (Broekstra 1981):

$$(36) \quad H_Y(X) = H(X, Y) - H(Y).$$

This quantity, $H_Y(X)$, is a measure of one's average uncertainty as to the state of X , given the knowledge about the state of Y . Its value ranges from 0, when Y completely determines X , to $H(X)$ for complete independence between X and Y . One may also note the well known fact that:

$$(37) \quad H_Y(X) \leq H(X),$$

to which an intuitive meaning may be attached. That is, knowledge about the state of Y will, in general, not increase the uncertainty existing about the state of X . In expression (37), equality holds if and only if X and Y are independent, while $H_Y(X)$ vanishes if and only if the state of X is completely determined once the state of Y is known.

As pointed out by Krippendorff (1979), whenever the

probabilities of joint events are the product of the probabilities of individual events, entropies are additive, and the lack of additivity signifies interdependence. Thus, if X and Y are independent, $H(X,Y) = H(X) + H(Y)$. The difference between $H(X,Y)$ (i.e., the actual entropy) and $H(X) + H(Y)$ (the maximal entropy) is then a measure of interdependence.

In developing information theory into a multivariable calculus, Garner and McGill (1956) and Ashby (1964) have generalized Shannon's (Shannon and Weaver 1949) binary notion of "association" into a multivariate expression known as "transmission" or amount of "transmitted information," denoted by T . For two variables X and Y , for instance, the degree of association between X and Y , denoted by $T(X:Y)$, may be expressed as the difference between the marginal and the joint variability. That is, transmission $T(X:Y)$ is given by:

$$\begin{aligned}
 (38) \quad T(X:Y) &= H_{\max}(X,Y) - H(X,Y) \\
 &= H(X) + H(Y) - H(X,Y) \\
 &= H(X) - H_Y(X).
 \end{aligned}$$

The above expressions measure the amount of statistical dependence (or constraint or relatedness) of a pair of variables X and Y . Transmission is symmetric, i.e., $T(X:Y) = T(Y:X)$, and non-negative. It is equal to zero if and only if the variables are statistically independent, and equal to the smaller of $\{H(X), H(Y)\}$ if and only if the knowledge of

the state of one variable completely reveals the state of the other variable. (Broekstra 1981).

Owing to this behavior of the transmissions with respect to its range, $0 \leq T(X:Y) \leq \text{Min}\{H(X), H(Y)\}$, its numerical value may be thought of as expressing the strength of the interdependence between X and Y. This measure is completely decomposable, which can be partitioned into variability between and variability within sets of variables, allowing for ANOVA-like analysis of variables (Garner and McGill 1956, Horowitz and Horowitz 1976).

For the multivariate case, expression (38) can easily be shown as (Garner and McGill 1956, Horowitz and Horowitz 1976):

$$(39) \quad T(A:B:C:\dots:X) = H(A) + H(B) + H(C) + \dots + H(X) - H(A, B, C, \dots, X).$$

Equation (39) represents the total amount of interdependence or constraint existing within the system as a whole. If this quantity is equal to zero, the variables are thought to be jointly independent. The maximum for equation (39) is defined as $H(i) - \max\{H(i)\}; i=A, B, C, \dots, X$.

Another important type of transmission function remains to be discussed. That is the information transmitted from one variable to another where a third is already known (or held constant). For instance, information shared by A and B when X is held constant is denoted by

$T_X(A:B)$. This measure can be calculated as (Garner and McGill 1956, Ashby 1964, Krippendorff 1979):

$$(40) \quad T_X(A:B) = T(A:B:X) - T(A:X) - T(B:X);$$

where $T_X(A:B)$ is the conditional transmissions which measures statistical dependence between the pair of variables A and B, while a third variable, X, is held constant. In general (Garner and McGill 1956, Ashby 1964, Krippendorff 1979):

$$(41) \quad T_X(A:B:C:\dots) = T(A:B:\dots:X) - T(A:X) - T(B:X) - T(C:X) - \dots$$

Also, the uncertainty of a variable in a multivariate system can be decomposed into the portion explained by other variables (i.e., transmissions between that variable and the remaining variables), and the portion not represented by any other variable (i.e., the uncertainty specific to the variable under consideration) (Abrahamse and van Bueren 1980, Krippendorff 1986). For instance, in a four variable system {WXYZ}, the following breakdown of the uncertainty for the variable W may result:

$$(42) \quad H(W) = T(W:X) + T_X(W:Y) + T_{XY}(W:Z) + H_{XYZ}(W)$$

In Equation (42), $T_X(W:Y)$ and $T_{XY}(W:Z)$ are conditional transmissions, which have been defined above, and the conditional entropy, $H_{XYZ}(W)$, is the amount of uncertainty remaining in variable W when the states of X, Y, and Z are known.

Equation (42) is useful in prioritizing the contributions to the reduction of the uncertainty in a criterion variable (e.g., W) by a number of explanatory variables (e.g., X , Y , and Z) through sequentially selecting the highest contributors. For instance, in choosing the highest contributor to the reduction of uncertainty in W , $\text{MAX}\{T(W:i), i = X, Y, \text{ or } Z\}$ is selected. When selecting the second highest contributor, $\text{MAX}\{T_i(W:j), j = X, Y, \text{ or } Z, j \neq i\}$ is selected, and so on. This operation first identifies the highest contributor, then, assuming that this variable is given, identifies the next highest contributor, thus eliminating the overlap between the two explanatory variables in reducing the uncertainty in the criterion variable. Therefore, each variable's incremental effect on uncertainty reduction of the criterion variable is measured in each successive step and the highest contributor is selected.

Another quantity which is useful in assessing different structural models (as will be addressed in detail in the next section) is the amount of "interaction." Interaction is a measure of the uniqueness of the relationship within a set of variables, attributed only to the n -way effects among them. For instance, the interaction between three variables, denoted by $Q(A,B,C)$, has a unique value, which may be defined in one of the following equivalent ways (Broekstra 1981):

$$\begin{aligned}
 (43) \quad Q(A,B,C) &= T_A(B:C) - T(B:C) \\
 &= T_B(A:C) - T(A:C) \\
 &= T_C(A:B) - T(A:B).
 \end{aligned}$$

Note that interaction term Q can be either positive or negative since transmissions and conditional transmissions are non-negative quantities.

This survey is voluntary and authorized under the Pacific Northwest Electric Power Planning and Conservation Act of 1980 (Public Law 96-501).

Spring 1983

**U.S. Department of Energy
Bonneville Power Administration**

[illegible]

ERRATA

Due to an oversight, the SPSS Data Definition (DD) Files for the NOAA Weather Data files were improperly specified. The DD File contained in the computer tape is formatted to read only one record per case. However, the actual data contains four records per case, each record containing the data for one of the four sets of heating degree days (HDD) and cooling degree days (CDD) base temperatures.

Please note that to read the data properly, the user must alter the format statement and the list of variables. Following is a copy of the DD file contained on the tape:

ID,CZ,HHDBASE,CDDBASE,HDD001 TO HDD020,CDD001 TO CDD020

(4F5.0,40F8.0)

This DD is used for both the electricity billing weather data and the gas billing weather data.

For electricity billing weather data, the DD file should read as follows:

EID1,ECD1,EHDD1B,ECDD1B,EHDD1001 to EHDD1020,ECDD1001 to ECDD1020,

EID2,ECD2,EHDD2B,ECDD2B,EHDD2001 to EHDD2020,ECDD2001 to ECDD2020,

EID3,ECD3,EHDD3B,ECDD3B,EHDD3001 to EHDD3020,ECDD3001 to ECDD3020,

EID4,ECD4,EHDD4B,ECDD4B,EHDD4001 to EHDD4020,ECDD4001 to ECDD4020,

(4F5.0,40F8.0/4F5.0,40F8.0/4F5.0,40F8.0/4F5.0,40F8.0)

For the gas billing weather data, the DD file should read as follows:

GID1,GCD1,GHDD1B,GCDD1B,GHDD1001 to GHDD1020,GCDD1001 to GCDD1020,

GID2,GCD2,GHDD2B,GCDD2B,GHDD2001 to GHDD2020,GCDD2001 to GCDD2020,

GID3,GCD3,GHDD3B,GCDD3B,GHDD3001 to GHDD3020,GCDD3001 to GCDD3020,

GID4,GCD4,GHDD4B,GCDD4B,GHDD4001 to GHDD4020,GCDD4001 to GCDD4020,

(4F5.0,40F8.0/4F5.0,40F8.0/4F5.0,40F8.0/4F5.0,40F8.0)

1a. Is there any apartment, guest house, or other type of separate living quarters served by the same electric meter you are connected with?

VAR

- [0] ☐ No
 [1] ☐ Yes
 [8] ☐ Don't Know

SKIP TO Q-2

003

IF "YES," ASK:

b. Do any of these living quarters have separate cooking facilities?

- [1] ☐ Yes
 [0] ☐ No
 [8] ☐ Don't Know

004

IF "NO" OR "DON'T KNOW," SAY:

c. When you answer the following questions, please include those living quarters and the occupants as if they were part of your household.
 GO TO Q-2

IF "YES," SAY:

d. Thank you. Our survey covers only those living quarters for which the electricity is separately metered. We appreciate your cooperation.

END THE INTERVIEW. RECORD "MASTER METER" ON CONTROL FORM.
 DON'T COMPLETE A QUESTIONNAIRE.

DATE: _____ (mnth) _____ (day) _____ AM
TIME INTERVIEW STARTED: _____ PM

2. Record type of dwelling. If in doubt, verify with respondent, using Exhibit #2.

[01] ☐ Mobile home or trailer—SKIP TO Q-5 (MH)

- [11] ☐ Single family house (SFH)

House or building with,

[22] ☐ 2 units (2U)

- [23] ☐ 3 units (3U)

- [24] ☐ 4 units (4U)

[31] ☐ Building with 5 or more units (5+)

[41] ☐ Other—Describe: OTH

IF "BUILDING WITH 5 OR MORE UNITS," ASK:

3. How many units does this building contain? _____ (Units)
SKIP TO Q-5

IF "SINGLE FAMILY" OR "BUILDING WITH 2 TO 4 UNITS," ASK:

4. Record the following. If in doubt, verify with respondent. House or building is . . .

[1] ☐ Detached

[2] ☐ Attached on one side

[3] ☐ Attached on two sides

[7] ☐ Not applicable

5. Do you or members of your household own your home or do you rent?

[1] ☐ Own or Buying—SKIP TO Q-7

[2] ☐ Rent

[3] ☐ Occupied without payment of rent

IF "RENT" OR "OCCUPIED WITHOUT RENT," ASK:

6. Which of the following are furnished as part of the house (apartment)?
 READ LIST

No	Yes	N/A
[0]	[1]	[7]

Refrigerator ☐ ☐ ☐

Stove top/burners . ☐ ☐ ☐

Oven ☐ ☐ ☐

Clothes Washer ... ☐ ☐ ☐

Clothes Dryer ☐ ☐ ☐

Dishwasher..... ☐ ☐ ☐

7. In what year did you first move into this house (apartment)?

_____ Actual year

[998] ☐ Don't Know

VAR
015

IF "DON'T KNOW," ASK:

8. Would you mind reviewing Exhibit #8 and indicating the category which you think best describes your estimate of the time when you first moved into this house (apartment)?

[01] ☐ (a) Before 1940

[06] ☐ (f) 1975-1978

[02] ☐ (b) 1940-1949

[07] ☐ (g) 1979-1981

[03] ☐ (c) 1950-1959

[08] ☐ (h) 1982-1983

[04] ☐ (d) 1960-1969

[98] ☐ Don't Know

[05] ☐ (e) 1970-1974

[97] ☐ Not Applicable

016

9. Did any of the current members of your household move into this house (apartment) before you moved in?

[0] ☐ No

[1] ☐ Yes

[8] ☐ Don't Know

SKIP TO Q-12

017

IF "YES," ASK:

10. In what year did this (these) member(s) of your household first move into this house (apartment)? _____ Actual year

[998] ☐ Don't Know

018

IF "DON'T KNOW," ASK:

11. Would you review Exhibit #8 and indicate the category which best describes your estimate of the time when this (these) member(s) of your household first moved into this house (apartment)?

[01] ☐ (a) Before 1940

[06] ☐ (f) 1975-1978

[02] ☐ (b) 1940-1949

[07] ☐ (g) 1979-1981

[03] ☐ (c) 1950-1959

[08] ☐ (h) 1982-1983

[04] ☐ (d) 1960-1969

[98] ☐ Don't Know

[05] ☐ (e) 1970-1974

[97] ☐ Not Applicable

019

IF "1981," "1982" OR "1983," ASK:

12. In which year and month did you (this/these member[s]) first move in? (Circle the applicable months.)

Year: ☐ 79-80 ☐ 81 ☐ 82 ☐ 83 ☐ Don't Know ☐ N/A
[80] [81] [82] [83] [98] [97]

020

Month: J F M A M JN JL A S O N D
[01] [02] [03] [04] [05] [06] [07] [08] [09] [10] [11] [12]

021

[98] ☐ Don't Know [97] ☐ Not Applicable

INTERVIEWER INSTRUCTIONS:

For Q-12 and for subsequent questions which refer to "date moved in," use date provided for household member who moved in first.

13. Was this house (building) newly constructed when you (member[s] of your household) first moved in? **VAR**

- [0] ☐ No
 [1] ☐ Yes **022**
 [8] ☐ Don't Know/Don't Remember

14. In what year was this house (building) constructed?

- _____ (actual year) **023**
 [998] ☐ Don't Know

IF R "DOES NOT KNOW," ASK:

15. Please review the list on Exhibit #8 and indicate the category which best describes your estimate of the time when this house (building) was constructed.

- [01] ☐ (a) Before 1940
 [02] ☐ (b) 1940-1949
 [03] ☐ (c) 1950-1959
 [04] ☐ (d) 1960-1969
 [05] ☐ (e) 1970-1974
 [06] ☐ (f) 1975-1978
 [07] ☐ (g) 1979-1981
 [08] ☐ (h) 1982-1983
 [98] ☐ Don't Know
 [97] ☐ Not Applicable **024**

16. Has this house (building) been substantially renovated or remodeled since you (member[s] of your household) first moved in? That is, has work costing at least \$5,000 been done on the house, and including such things as the addition of a new room; the rearrangement of internal walls; the installation of new windows or complete overhaul of existing windows? Redecoration of the house is not included—for example new paint or wallpaper, new carpets or floors, or new cabinets.

- [0] ☐ No
 [1] ☐ Yes **025**
 [8] ☐ Don't Know

17. Exhibit #17 contains a list of different types of rooms which you may have in your home. Please tell me how many of each of these types of rooms there are in your home.

		Q-17	Q-18	
(a) Living room	026	<input type="text"/>	<input type="text"/>	027
(b) Dining room	028	<input type="text"/>	<input type="text"/>	029
(c) Kitchen, <u>with</u> eating area	030	<input type="text"/>	<input type="text"/>	031
(d) Kitchen, <u>without</u> eating area	032	<input type="text"/>	<input type="text"/>	033
(e) Bedroom	034	<input type="text"/>	<input type="text"/>	035
(f) Den, Study, Library, Sewing room	036	<input type="text"/>	<input type="text"/>	037
(g) Family room, Recreation room	038	<input type="text"/>	<input type="text"/>	039
(h) Utility or Work room	040	<input type="text"/>	<input type="text"/>	041
(i) Other small room (80 Sq. Ft. or less)	042	<input type="text"/>	<input type="text"/>	043
(j) Other large room (more than 80 Sq. Ft.)	044	<input type="text"/>	<input type="text"/>	045
TOTAL NUMBER OF ROOMS	046	<input type="text"/>	<input type="text"/>	047

18a. Are any of these rooms not heated during most months when heating is necessary?

No ☐ [0]

Yes ☐ [1]

048

IF "YES," ASK:

18b. Using Exhibit #17, please indicate how many of each of these types of rooms are not heated.

INTERVIEWER INSTRUCTIONS:

Q-17: Do not count bathrooms, entryways, hallways, unheated porches, unfinished attics, or unfinished basements. Count living room/dining room combinations (no partitioning wall between) as a single room and code as "Living room." Probe to determine whether there is an attic and, if so, whether it is heated.

Q-18: Count a room as not heated if the door is closed and the heat is shut off or vents are closed. The room may still be used (for sleeping for example), if the door is closed and the heat is off while the room is in use. If there is an attic, probe to determine whether it is heated.

VAR

19. Please review the list on Exhibit # 19 and indicate which of the statements best describes how you feel about the energy efficiency of your home?

- [1] ☐ (a) The home is about as energy efficient as it can be
 [2] ☐ (b) A little improvement can be made
 [3] ☐ (c) Moderate improvement can be made
 [4] ☐ (d) A lot of improvement can be made
 [8] ☐ Don't Know
 [9] ☐ Refused

049

20. I am going to read you a series of items. Each item refers to an issue which some people may consider to be a problem in their state. We would like to know your thoughts about how serious each of these issues is today in your state. Exhibit #20 contains a list of the answers you may choose from. As I read each item, please tell me whether you consider the item to be not serious at all, slightly serious, moderately serious or very serious.

	Not Serious [1]	Slightly Serious [2]	Moderately Serious [3]	Very Serious [4]	Don't Know [8]	
(a) Environmental pollution in your state .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	050
(b) Cost of energy in your state.....	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	051
(c) Unemployment in your state	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	052
(d) Scarcity of energy in your state	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	053
(e) Inflation in your state	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	054
(f) Crime in your state	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	055

IF "COST OF ENERGY" OR "SCARCITY OF ENERGY" RATED
 "MODERATELY SERIOUS" OR "VERY SERIOUS," ASK:

21. Now please turn to Exhibit #21 and, for each of the forms of energy listed, please indicate whether (cost/scarcity) is a problem in your state.
 READ LIST

		COST			SCARCITY			
		No [0]	Yes [1]	DK [8]	No [0]	Yes [1]	DK [8]	
(a) Electricity	056	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	057
(b) Natural Gas	058	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	059
(c) Heating Oil	060	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	061
(d) Gasoline								
(for transportation) ...	062	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	063
(e) Wood	064	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	065

22. If you would please turn to Exhibit #22 you will find a list of statements. I am going to read each of these statements to you. For each statement, please tell me how strongly you agree or disagree with the statement. At the bottom of Exhibit #22 there is a list of the answers you may choose from. Let us begin with the first statement.

VAR

	Strongly Agree	Somewhat Agree	Undecided	Somewhat Disagree	Strongly Disagree	
(a) In the winter, I find it difficult to be comfortable when the temperature in my home is 68°F or less.	<input type="checkbox"/> (1)	<input type="checkbox"/> (2)	<input type="checkbox"/> (3)	<input type="checkbox"/> (4)	<input type="checkbox"/> (5)	066
(b) Reducing the temperature of the water heater from 140°F to 120°F saves enough energy to make it worth doing. ...	<input type="checkbox"/> (1)	<input type="checkbox"/> (2)	<input type="checkbox"/> (3)	<input type="checkbox"/> (4)	<input type="checkbox"/> (5)	067
(c) The main reason to conserve energy is to save money.	<input type="checkbox"/> (1)	<input type="checkbox"/> (2)	<input type="checkbox"/> (3)	<input type="checkbox"/> (4)	<input type="checkbox"/> (5)	068
(d) During the winter, when no one will be home for 2 hours or more, turning the temperature down is worthwhile.	<input type="checkbox"/> (1)	<input type="checkbox"/> (2)	<input type="checkbox"/> (3)	<input type="checkbox"/> (4)	<input type="checkbox"/> (5)	069
(e) It's hard for me to get around to making my home more energy efficient.	<input type="checkbox"/> (1)	<input type="checkbox"/> (2)	<input type="checkbox"/> (3)	<input type="checkbox"/> (4)	<input type="checkbox"/> (5)	070
(f) People have a right to use as much energy as they want and can pay for.	<input type="checkbox"/> (1)	<input type="checkbox"/> (2)	<input type="checkbox"/> (3)	<input type="checkbox"/> (4)	<input type="checkbox"/> (5)	071
(g) The price I first pay for the appliance is more important to me than the energy savings.	<input type="checkbox"/> (1)	<input type="checkbox"/> (2)	<input type="checkbox"/> (3)	<input type="checkbox"/> (4)	<input type="checkbox"/> (5)	072
(h) To conserve enough energy to make a difference in my bills I would have to change my lifestyle.	<input type="checkbox"/> (1)	<input type="checkbox"/> (2)	<input type="checkbox"/> (3)	<input type="checkbox"/> (4)	<input type="checkbox"/> (5)	073

23. Please review the list on Exhibit #23 and, for each item, indicate whether you have heard or read anything about the item. READ LIST

	No	Q-23 Yes	N/A	No	Q-24 Yes	N/A	
(a) Federal income tax credit for home weatherization 074	<input type="checkbox"/> (0)	<input type="checkbox"/> (1)	<input type="checkbox"/> (7)	<input type="checkbox"/> (0)	<input type="checkbox"/> (1)	<input type="checkbox"/> (7)	075
(b) State tax benefit for home weatherization 076	<input type="checkbox"/> (0)	<input type="checkbox"/> (1)	<input type="checkbox"/> (7)	<input type="checkbox"/> (0)	<input type="checkbox"/> (1)	<input type="checkbox"/> (7)	077
(c) Low or no interest loan from electric or gas utility or fuel oil supplier for home weatherization 078	<input type="checkbox"/> (0)	<input type="checkbox"/> (1)	<input type="checkbox"/> (7)	<input type="checkbox"/> (0)	<input type="checkbox"/> (1)	<input type="checkbox"/> (7)	079

FOR EACH "YES," ASK:

24. Have you made use of _____ ? _____

25. To the best of your knowledge, has an energy professional made a physical inspection of your home, including measurement of insulation levels, and advised you on ways to make your home more energy efficient?

VAR

- [1] ☐ Yes
 [0] ☐ No
 [8] ☐ Don't Know

080

IF "NO" OR "DON'T KNOW," ASK:

26. From the list on Exhibit #26, please indicate the category which best describes how likely it is that you would agree to have an energy professional visit your home, at no cost to you, and advise you on ways to make your home more energy efficient.

- [1] ☐ Not at all likely
 [2] ☐ Slight likelihood
 [3] ☐ Moderate likelihood
 [4] ☐ Very likely—certain or almost certain
 [5] ☐ Requested, but not yet received
 [8] ☐ Don't Know
 [7] ☐ Not Applicable

081

SKIP TO Q-31(B)

IF "YES," ASK:

27. When was your home last inspected by an energy professional?

[If month is DK, Code "00"]

Month _____ Year _____

082

- [98] ☐ Don't Know—Month & Year

083

- [97] ☐ Not Applicable

IF RESPONDENT "DOES NOT KNOW," ASK:

28. Would you mind turning to Exhibit #28 and indicating your best estimate of the time when your home was last inspected by an energy professional?

- [1] ☐ Before 1/1/78
 [2] ☐ Between 1/1/78 and 12/31/79
 [3] ☐ Between 1/1/80 and 12/31/81
 [4] ☐ After 1/1/82
 [8] ☐ Don't Know
 [7] ☐ Not Applicable

084

29. Was the inspection conducted by a representative of your electric utility, your gas utility, a heating oil company or by a private contractor?

- [1] ☐ Electric Utility
 [2] ☐ Gas Utility
 [3] ☐ Heating Oil Company
 [4] ☐ Private Contractor
 [8] ☐ Don't Know
 [7] ☐ Not Applicable

085

30a. Exhibit #30/31 contains a list of energy saving measures. Please review the list and, for each measure, indicate whether it was included in the energy professional's recommendations of measures to be taken to improve the energy efficiency of your home. READ EACH ITEM

Q-30a				Q-30b				Q31a/b				
No	Yes	DK		No	Yes	DK		No	Yes	DK		
[0]	[1]	[8]		[0]	[1]	[8]		[0]	[1]	[8]		
086	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(a) Outside wall insulation	087	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	088
089	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(b) Roof/ceiling insulation	090	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	091
092	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(c) Floor insulation	093	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	094
095	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(d) Insulation of heating ducts	096	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	097
098	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(e) Storm doors	099	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	100
101	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(f) Storm windows/insulated glass	102	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	103
104	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(g) Plastic over windows	105	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	106
107	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(h) Wrap water heater	108	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	109
110	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(i) Insulation of hot water pipes	111	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	112
113	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(j) Weatherstripping	114	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	115
116	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(k) Caulking	117	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	118
119	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(l) Reduced water heater temperature	120	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	121
122	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(m) Reduced home heating temperature	123	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	124
125	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(n) Automatic setback thermostat	126	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	127
128	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(o) Other (Describe: _____)	129	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	130

30b. Using the same list, would you please indicate which actions you have taken since the inspection was completed?

IF INSPECTION COMPLETED SINCE "MAY 1980," ASK:

31a. Now, would you please indicate which actions, if any, you had taken before the inspection was completed, but within the last 3 years, that is, since May 1980? _____

IF "NO ENERGY INSPECTION" (Q-25), ASK:

31b. Would you please review the list on Exhibit #30/31 and indicate which actions you have taken in the last three years—that is, since May 1980? _____

VAR

32. To the best of your knowledge, which of the categories on Exhibit #32 best describes the percentage of the outside walls of your home that currently contain insulation?

- [0] ☐ (a) None—SKIP TO Q-34
 [1] ☐ (b) Some, but less than 10%
 [2] ☐ (c) 10-39% (about one quarter)
 [3] ☐ (d) 40-59% (about half)
 [4] ☐ (e) 60-89% (about three quarters)
 [5] ☐ (f) 90-100% (all, nearly all)
 [8] ☐ (g) Don't Know

131

IF AT LEAST "SOME," ASK:

33. Have you installed or added any insulation in the outside walls since May 1, 1982?

- [0] ☐ No
 [1] ☐ Yes, installed/added
 [8] ☐ Don't Know
 [7] ☐ Not Applicable

132

34. To the best of your knowledge, which of the categories on Exhibit #32 best describes the percentage of the roof, ceiling, or attic of your home that currently contains insulation?

- [0] ☐ (a) None—SKIP TO Q-36
 [1] ☐ (b) Some, but less than 10%
 [2] ☐ (c) 10-39% (about one quarter)
 [3] ☐ (d) 40-59% (about half)
 [4] ☐ (e) 60-89% (about three quarters)
 [5] ☐ (f) 90-100% (all, nearly all)
 [8] ☐ (g) Don't Know

133

IF AT LEAST "SOME," ASK:

35. Have you installed or added any insulation in the roof, ceiling, or attic since May 1, 1982?

- [0] ☐ No
 [1] ☐ Yes, installed/added
 [8] ☐ Don't Know
 [7] ☐ Not Applicable

134

IF "2 OR MORE UNITS," SKIP TO Q-42.

IF "ONE-FAMILY HOME" OR "MOBILE HOME," ASK:

36. Please turn to Exhibit #36 and indicate the categories which best describe your home. Please indicate all that apply.

Yes
[1]

No
[0]

☒

Basement

☐

Crawl Space—enclosed

☐

Crawl Space—open to outside

☐

Concrete Slab

☐

Other: (Describe) _____

135

136

137

138

139

IF "BASEMENT," ASK:

37. Please turn to Exhibit #37 and indicate the category which best describes the percentage of the basement which is heated. That is, what percentage is a comfortable place to rest, read, study or watch television, year-round?

[0] ☐ (a) None

[1] ☐ (b) Some, but less than 10%

[2] ☐ (c) 10-39% (about one quarter)

[3] ☐ (d) 40-59% (about half)

[4] ☐ (e) 60-89% (about three quarters)

[5] ☐ (f) 90-100% (all, nearly all)

[8] ☐ (g) Don't Know

[7] ☐ Not Applicable

140

VAR

IF "CRAWL SPACE" (Q-36) OR IF LESS THAN 90% OF BASEMENT IS HEATED, (Q-37)

ASK:

38. To the best of your knowledge, which of the categories on Exhibit #37 best describes the percentage of the floor area above the unheated portion of the basement (crawl space) that is currently insulated?

[0] ☐ (a) None—SKIP TO Q-40

[1] ☐ (b) Some, but less than 10%

[2] ☐ (c) 10-39% (about one quarter)

[3] ☐ (d) 40-59% (about half)

[4] ☐ (e) 60-89% (about three quarters)

[5] ☐ (f) 90-100% (all, nearly all)

[8] ☐ (g) Don't Know

[7] ☐ Not Applicable

IF AT LEAST "SOME," ASK:

39. Have you installed or added any insulation in the floor area above the unheated basement (crawl space) since May 1, 1982?

[0] ☐ No

[1] ☐ Yes, installed/added

[8] ☐ Don't Know

[7] ☐ Not Applicable

141

142

VAR

IF BETWEEN "SOME" AND "ALL" OF THE BASEMENT IS HEATED, (Q-37) ASK:

40. To the best of your knowledge, which of the categories on Exhibit #37 best describes the percentage of the walls (of the heated portion) of the basement that currently contain insulation?

- [0] ☐ (a) None—SKIP TO Q-42
- [1] ☐ (b) Some, but less than 10%
- [2] ☐ (c) 10-39% (about one quarter)
- [3] ☐ (d) 40-59% (about half)
- [4] ☐ (e) 60-89% (about three quarters)
- [5] ☐ (f) 90-100% (all, nearly all)
- [8] ☐ (g) Don't Know
- [7] ☐ Not Applicable

143

IF AT LEAST "SOME," ASK:

41. Have you installed or added any insulation to the basement walls since May 1, 1982?

- [0] ☐ No
- [1] ☐ Yes, installed/added
- [8] ☐ Don't Know
- [7] ☐ Not Applicable

144

ASK EVERYONE:

42. Excluding sliding glass doors, how many doors are there in your home that lead directly from a heated area to the outside or to an unheated porch or unheated garage?

_____ Number doors
[00] ☐ None—SKIP TO Q-44

145

INTERVIEWER INSTRUCTIONS:

Do not include doors to unheated basements or unheated attics.
Do not include sliding glass doors in Q-42 and 43.

IF "ONE OR MORE DOORS," ASK:

43. How many of these doors have storm doors installed during the months when heating is required?

_____ Number storm doors
[00] ☐ None
[97] ☐ Not applicable

146

44. How many sliding glass doors are there in your home that lead directly from a heated area to the outside or to an unheated porch or unheated garage?

_____ Number sliding glass doors
[00] ☐ None—SKIP TO Q-46

147

IF "ONE OR MORE SLIDING GLASS DOORS," ASK:

45. How many of these sliding glass doors contain insulating glass or have storm doors installed during the months when heating is required?

_____ Number insulated glass
or storm doors
[00] ☐ None
[97] ☐ Not applicable

148

46. Please review Exhibit #46 and indicate which of the categories on the list apply to the windows in your home during the months when heating is required. READ LIST VAR

Yes [1]	No [0]	DK [8]		
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(a)	Insulated or thermopane glass 149
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(b)	Storm windows 150
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(c)	Plastic over windows 151
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	(d)	Other (Describe: _____) 152

IF "YES," TO ONE OR MORE, ASK:

47. Now would you please look at Exhibit #47 and indicate the category which best describes the percentage of windows which have...
READ EACH ITEM

	None (a)	Less Than 10% (b)	10-39 % About 1/4 (c)	40-59 % About 1/2 (d)	60-89 % About 3/4 (e)	90-100% All/ Nearly All (f)	DK (g)	
(a) Insulated/thermopane glass.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	153
(b) Storm windows	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	154
(c) Plastic over windows.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	155
(d) Other	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	156
	[0]	[1]	[2]	[3]	[4]	[5]	[8]	

Now I would like to ask you a few questions about how you heat your home.

VAR

48. Would you please turn to Exhibit #48 and indicate which one of the fuels listed is used most of the time to heat your home?

- [01] ☐ (a) Wood or prestologs 157
- [02] ☐ (b) Electricity
- [03] ☐ (c) Natural gas from underground pipes
- [04] ☐ (d) Fuel oil
- [05] ☐ (e) Propane, LPG (Bottled or tank gas)—SKIP TO Q-53
- [06] ☐ (f) Kerosene
- [07] ☐ (g) Coal
- [08] ☐ (h) Solar collectors — SKIP TO Q-53
- [09] ☐ (i) Geothermal
- [20] ☐ (j) Other (Specify: _____)—SKIP TO Q-53
- [00] ☐ (k) None (not heated)—SKIP TO Q-71
- [98] ☐ Don't Know—SKIP TO Q-53

IF "NATURAL GAS," "FUEL OIL," OR "KEROSENE," ASK:

49. Please review Exhibit #49 and indicate which one of the types of equipment listed is used most of the time to heat your home.

- [1] ☐ (a) Central warm air or forced air furnace
- [2] ☐ (b) Steam or hot water system with radiators, convectors or pipes running through a slab floor
- [3] ☐ (c) Unit(s) which are permanently installed in the wall, floor or ceiling 158
- [4] ☐ (d) Portable units
- [5] ☐ (e) Other (Describe: _____)
- [8] ☐ (f) Don't Know
- [7] ☐ Not Applicable

IF "ELECTRICITY," (Q-48) ASK:

50. Please review Exhibit #50 and indicate which one of the types of equipment listed is used most of the time to heat your home.

- [1] ☐ (a) Heat pump
- [2] ☐ (b) Central warm air or forced air furnace
- [3] ☐ (c) Baseboard or wall units
- [4] ☐ (d) Radiant panels in floor, ceiling or walls
- [5] ☐ (e) Portable room heater(s)
- [6] ☐ (f) Other (Describe: _____)
- [8] ☐ (g) Don't Know
- [7] ☐ Not Applicable

159

IF "WOOD," (Q-48) ASK:

51. Please review Exhibit #51 and indicate which one of the types of equipment listed is used most of the time to heat your home.

- [1] ☐ (a) Free-standing stove
- [2] ☐ (b) Fireplace(s)
- [3] ☐ (c) Central furnace with ducts to individual rooms
- [4] ☐ (d) Other (Describe: _____)
- [8] ☐ (e) Don't Know
- [7] ☐ Not Applicable

160

IF "FIREPLACE(S)," ASK FIRST ABOUT FIREPLACE USED MOST:

52. Are any of the items listed on Exhibit #52 installed in your fireplace?

- | | FP#1 | FP#2 | |
|---------|--------------------------|--------------------------|-----------------------------|
| [1] | <input type="checkbox"/> | <input type="checkbox"/> | (a) Stove insert |
| [2] | <input type="checkbox"/> | <input type="checkbox"/> | (b) Heatolator |
| 161 [3] | <input type="checkbox"/> | <input type="checkbox"/> | (c) Glass doors on front |
| [4] | <input type="checkbox"/> | <input type="checkbox"/> | (d) Other (Describe: _____) |
| [0] | <input type="checkbox"/> | <input type="checkbox"/> | (e) No additional equipment |
| [7] | <input type="checkbox"/> | <input type="checkbox"/> | Not Applicable |

162

IF 2 OR MORE HOUSING UNITS IN BUILDING, ASK:

53. Is your home heated by a central system for your building (or group of buildings) or is the main heating equipment for your living quarters only?

- [1] ☐ Central system for building(s)
 [2] ☐ Main heat equipment for these living quarters only
 [8] ☐ Don't Know
 [7] ☐ Not Applicable

VAR

163

ASK EVERYONE:

54. Since September 1981 have you changed the fuel type you use most of the time to heat this house (apartment)?

- [0] ☐ No
 [1] ☐ Yes
 [7] ☐ No Fuel Used
 [8] ☐ Don't Know

164

IF "YES," ASK:

55. In what month and year did you make this change?
 [If month is DK, code "00."]

_____ Month
 _____ Year

- [98] ☐ Don't Know—Month and Year
 [97] ☐ Not Applicable

165

166

56. Please look at Exhibit #56 and indicate which fuel was used most of the time to heat this home before September 1981?

- [01] ☐ (a) Wood or prestologs
 [02] ☐ (b) Electricity
 [03] ☐ (c) Natural gas from underground pipes
 [04] ☐ (d) Fuel oil
 [05] ☐ (e) Propane, LPG (Bottled or tank gas)
 [06] ☐ (f) Kerosene
 [07] ☐ (g) Coal
 [08] ☐ (h) Solar collectors
 [09] ☐ (i) Geothermal
 [20] ☐ (j) Other (Describe: _____)
 [00] ☐ (k) None (not heated)
 [98] ☐ Don't Know
 [97] ☐ Not Applicable

167

ASK EVERYONE:

VAR

57. Was the heating equipment you currently use most of the time already installed in this home when you (member[s] of your household) first moved in?

[0] ☐ No[1] ☐ Yes

168

[8] ☐ Don't Know

58. In what year was the heating equipment you now use most of the time installed in this home?

_____ (Actual Year)

169

[998] ☐ Don't Know

IF "DON'T KNOW," ASK:

59. Would you mind reviewing Exhibit #59 and indicating the category which best describes your estimate of the time when the heating equipment you now use most was installed?

[01] ☐ (a) Before 1930[02] ☐ (b) 1930-1939[03] ☐ (c) 1940-1949[04] ☐ (d) 1950-1959[05] ☐ (e) 1960-1964[06] ☐ (f) 1965-1969[07] ☐ (g) 1970-1974[08] ☐ (h) 1975-1978[09] ☐ (i) 1979-1981[10] ☐ (j) Since 1/1/82[98] ☐ Don't Know[97] ☐ Not Applicable

170

ASK EVERYONE:

VAR

60. You have already mentioned the fuel which you use most of the time to heat your home.
Do you use any other fuels to heat your home in addition to the fuel you use most of the time?

[0] ☐ No—SKIP TO Q-66

[1] ☐ Yes

171

IF "YES," ASK:

61. Please look at Exhibit #61 and indicate which fuels you use for heating this home in addition to the fuel used most of the time.
Please indicate all of the fuels used.

Yes
[1]

No
[0]

☐

(a) Wood or prestologs

172

☐

(b) Electricity

173

☐

(c) Natural gas from underground pipes

174

☐

(d) Fuel oil

175

☐

(e) Propane, LPG (Bottled or tank gas)

176

☐

(f) Kerosene

177

☐

(g) Coal

178

☐

(h) Solar collectors

179

☐

(i) Geothermal

180

☐

(j) Other (Describe: _____)

181

IF "NATURAL GAS," "FUEL OIL," OR "KEROSENE," ASK:

62. Please review Exhibit #62 and indicate which one of the types of equipment listed is used to provide additional heat for your home.

[1] ☐ (a) Central warm air or forced air furnace

[2] ☐ (b) Steam or hot water system with radiators, convectors or pipes running through a slab floor

[3] ☐ (c) Unit(s) which are permanently installed in the wall, floor or ceiling

[4] ☐ (d) Portable units

[5] ☐ (e) Other (Describe: _____)

[8] ☐ Don't Know

[7] ☐ Not Applicable

182

VAR

↓

IF "ELECTRICITY," (Q-61) ASK:

63. Please review Exhibit #63 and indicate which one of the types of equipment listed is used to provide additional heat for your home.

- [1] ☐ (a) Heat pump
- [2] ☐ (b) Central warm air or forced air furnace
- [3] ☐ (c) Baseboard or wall units
- [4] ☐ (d) Radiant panels in floor, ceiling or walls
- [5] ☐ (e) Portable room heater(s)
- [6] ☐ (f) Other (Describe: _____)
- [8] ☐ (g) Don't Know
- [7] ☐ Not Applicable

183

↓

IF "WOOD," (Q-61) ASK:

64. Please review Exhibit #64 and indicate which one of the types of equipment listed is used to provide additional heat for your home.

- [1] ☐ (a) Free-standing stove
- [2] ☐ (b) Fireplace(s)
- [3] ☐ (c) Central furnace with ducts to individual rooms
- [4] ☐ (d) Other (Describe: _____)
- [8] ☐ (e) Don't Know
- [7] ☐ Not Applicable

184

↓

IF "FIREPLACE(S)," ASK FIRST ABOUT FIREPLACE USED MOST:

65. Are any of the items listed on Exhibit #65 installed in your fireplace(s)?

- | | FP#1 | FP#2 | |
|----------------------------------|--------------------------|-----------------------------|--|
| [1] <input type="checkbox"/> | <input type="checkbox"/> | (a) Stove insert | |
| [2] <input type="checkbox"/> | <input type="checkbox"/> | (b) Heatolator | |
| 185 [3] <input type="checkbox"/> | <input type="checkbox"/> | (c) Glass doors on front | |
| [4] <input type="checkbox"/> | <input type="checkbox"/> | (d) Other (Describe: _____) | |
| [0] <input type="checkbox"/> | <input type="checkbox"/> | (e) No additional equipment | |
| [7] <input type="checkbox"/> | <input type="checkbox"/> | Not Applicable | |

186

ASK EVERYONE:

VAR

66. Have you burned any wood or prestologs in your home in the past 12 months?

[Probe for type.]

[0] ☐ No—SKIP TO Q-70[1] ☐ Yes, wood[2] ☐ Yes, both wood and prestologs[3] ☐ Yes, prestologs

187

IF "PRESTOLOGS," ASK:

67. Please turn to Exhibit #67 and indicate the category which best describes your estimate of the number of logs which you burned in the last 12 months.

[1] ☐ (a) Less than 10[2] ☐ (b) 10 to 50[3] ☐ (c) 51 to 100[4] ☐ (d) 101 to 150[5] ☐ (e) 151 to 250[6] ☐ (f) More than 250[8] ☐ Don't Know[7] ☐ Not Applicable

188

IF "WOOD," ASK:

68. Using the illustration on Exhibit #68, would you please estimate how many cords of wood you burned in the past 12 months?

_____ Number of cords

[9998] ☐ Don't Know[9997] ☐ Not Applicable

189

INTERVIEWER INSTRUCTIONS:

Probe for respondent's best estimate of number of cords burned.
Record answer to nearest fraction of a cord—.25, 1.00, 1.50, 3.00, 6.00, 12.00.

69. Now please turn to Exhibit #69 and indicate the category that best describes your estimate of the percentage of this wood that you purchased.

[0] ☐ (a) None[1] ☐ (b) Some, but less than 10%[2] ☐ (c) 10-39% (about one quarter)[3] ☐ (d) 40-59% (about half)[4] ☐ (e) 60-89% (about three quarters)[5] ☐ (f) 90-100% (all, nearly all)[8] ☐ Don't Know[7] ☐ Not Applicable

190

ASK EVERYONE:

70. Do you use a thermostat, radiator valve, on-off switch or some other device to control the temperature in your home most of the time during the months when heating is required?

[0] ☐ No Control Device

[1] ☐ On-Off Switch

[2] ☐ Radiator Valve

191

[3] ☐ Thermostat

[4] ☐ Other (Describe: _____)

[8] ☐ Don't Know

IF "THERMOSTAT," ASK:

71. Is this an automatic clock thermostat or does the thermostat have a timer?

[0] ☐ No

[1] ☐ Yes, clock

[2] ☐ Yes, timer

192

[8] ☐ Don't Know

[7] ☐ Not Applicable

INTERVIEWER INSTRUCTIONS:

"CLOCK THERMOSTAT" = Automatically changes the temperature at pre-set times depending on time of day.

"TIMER THERMOSTAT" = Must be set each time it is used.

VAR

72. Now I'm going to ask you a few questions about the temperature you usually keep your home during the months when heating is required. Just your estimate or a range of temperatures will be fine. Please also tell me if you turn the heat off.

- (a) When someone is at home during the day or evening in the heating months?
(SEE INSTRUCTION BELOW.)

_____ Degrees Fahrenheit 193
194 _____ to _____ Degrees Fahrenheit (range) 195
[97] ☐ Heat turned off
[98] ☐ Don't Know

- (b) When no one is at home during the day or evening in the heating months?
(SEE INSTRUCTION BELOW.)

_____ Degrees Fahrenheit 196
197 _____ to _____ Degrees Fahrenheit (range) 198
[97] ☐ Heat turned off
[98] ☐ Don't Know

- (c) During sleeping hours in the heating months?
(SEE INSTRUCTION BELOW.)

_____ Degrees Fahrenheit 199
200 _____ to _____ Degrees Fahrenheit (range) 201
[97] ☐ Heat turned off
[98] ☐ Don't Know

INTERVIEWER INSTRUCTIONS:

If respondent keeps different sections of the house at different temperatures, we want to know the temperature in the part of the house where the people are. If, for example, the heat is turned off upstairs during the day because the family is downstairs, we want the downstairs temperature.
If respondent doesn't know temperature, but does know thermostat setting, record thermostat setting. Otherwise, probe for best estimate.

73. During the months when heating is required, is someone usually home and the house heated at least half of the time on weekdays between the hours of 9:00 a.m. and 5:00 p.m.?

VAR

[0] ☐ No

[1] ☐ Yes

202

[9] ☐ Refused

74. In order to help us better understand how you use energy in your home, would you mind telling us whether, since September 1981, there have been periods of at least one week when you did not use any of your appliances, heating or cooling equipment, or any other large equipment, because you were away on vacation, on business, were hospitalized, or for some other reason?

[0] ☐ No

[1] ☐ Yes

203

[8] ☐ Don't Know/Don't Remember

[9] ☐ Refused

IF "YES," ASK:

75. Please turn to Exhibit #75. During what month or months was this?
PROBE FOR ALL INSTANCES.

	No [0]	Yes [1]		No [0]	Yes [1]	
204	<input type="checkbox"/>	<input type="checkbox"/>	September 1981	<input type="checkbox"/>	<input type="checkbox"/>	January 1983 220
205	<input type="checkbox"/>	<input type="checkbox"/>	October	<input type="checkbox"/>	<input type="checkbox"/>	February 221
206	<input type="checkbox"/>	<input type="checkbox"/>	November	<input type="checkbox"/>	<input type="checkbox"/>	March 222
207	<input type="checkbox"/>	<input type="checkbox"/>	December	<input type="checkbox"/>	<input type="checkbox"/>	April 223
				<input type="checkbox"/>	<input type="checkbox"/>	May 224
208	<input type="checkbox"/>	<input type="checkbox"/>	January 1982			
209	<input type="checkbox"/>	<input type="checkbox"/>	February			
210	<input type="checkbox"/>	<input type="checkbox"/>	March	<input type="checkbox"/>	<input type="checkbox"/>	Don't Know/ Don't Remember 225
211	<input type="checkbox"/>	<input type="checkbox"/>	April			
212	<input type="checkbox"/>	<input type="checkbox"/>	May	<input type="checkbox"/>	<input type="checkbox"/>	Refused 226
213	<input type="checkbox"/>	<input type="checkbox"/>	June	<input type="checkbox"/>	<input type="checkbox"/>	Not Applicable 227
214	<input type="checkbox"/>	<input type="checkbox"/>	July			
215	<input type="checkbox"/>	<input type="checkbox"/>	August	[0] No	[1] Yes	
216	<input type="checkbox"/>	<input type="checkbox"/>	September			
217	<input type="checkbox"/>	<input type="checkbox"/>	October			
218	<input type="checkbox"/>	<input type="checkbox"/>	November			
219	<input type="checkbox"/>	<input type="checkbox"/>	December			
	[0] No	[1] Yes				

VAR

76. Now I would like to ask you some questions about how your water is heated.
Please turn to Exhibit #76 and indicate which fuel is used most for heating water in your home.

- [01] ☐ (a) Wood
- [02] ☐ (b) Electricity
- [03] ☐ (c) Natural gas from underground pipes
- [04] ☐ (d) Fuel oil
- [05] ☐ (e) Propane, LPG (Bottled or tank gas)
- [06] ☐ (f) Kerosene
- [08] ☐ (g) Solar collectors
- [10] ☐ (h) Heat pump
- [20] ☐ (i) Other (Describe: _____)
- [00] ☐ (j) None—no hot water—SKIP TO Q-91
- [98] ☐ Don't Know

228

IF 2 OR MORE HOUSING UNITS IN BUILDING, ASK:

77. Is your hot water supplied by a central system for your building
(or group of buildings) or is the water heater for your living quarters only?

- [1] ☐ Central system for building(s)
- [2] ☐ For these living quarters only
- [8] ☐ Don't Know
- [7] ☐ Not Applicable

229

78. Was the water heating equipment you currently use most already installed in your home
when you (member[s]) of your household first moved in?

- [0] ☐ No
- [1] ☐ Yes
- [8] ☐ Don't Know

230

INTERVIEWER INSTRUCTIONS:

Installation of a new hot water tank is considered
"new water heating equipment."
Installation of a new burner only is not considered
"new water heating equipment."

79. In what year was the water heating equipment you currently use most installed in your home?

VAR

_____ (Actual Year)

[998] ☐ Don't Know

231

IF "DON'T KNOW," ASK:

80. Would you mind looking at Exhibit #80 and indicating which of the categories best describes the time when the water heating equipment was installed?

[01] ☐ (a) Before 1960

[06] ☐ (f) 1979-1981

[02] ☐ (b) 1960-1964

[07] ☐ (g) Since 1/1/82

[03] ☐ (c) 1965-1969

[98] ☐ Don't Know

232

[04] ☐ (d) 1970-1974

[97] ☐ Not Applicable

[05] ☐ (e) 1975-1978

81. Is the water heater located in an area that is normally heated—or an area normally not heated?

[1] ☐ Heated area

[2] ☐ Not heated area

233

[8] ☐ Don't Know

INTERVIEWER INSTRUCTIONS:

An area is "heated" if it is a comfortable place to rest, read, study or watch television, year-round. An area is "heated" if it is a small area that is surrounded by heated space—for example, a closet.

82. Does that water heater have an extra layer of insulation wrapped around the outside of the tank—in the form of a mat or jacket or blanket?

[0] ☐ No—SKIP TO Q-84

[1] ☐ Yes

[2] ☐ Requested, but not yet installed

234

[8] ☐ Don't Know—SKIP TO Q-84

IF "YES," ASK:

83. Was the wrap (will the wrap be) provided by your fuel supplier (utility), did you purchase it yourself, or was it obtained from some other source?

[1] ☐ Supplier

[2] ☐ Self

[3] ☐ Other

235

[8] ☐ Don't Know

84. In addition to your main fuel, do you use any other fuel for heating water?

VAR

[0] ☐ No—SKIP TO Q-86

[1] ☐ Yes

236

IF "YES," ASK:

85. Please indicate what this additional fuel is, using Exhibit #85.

[01] ☐ (a) Wood

[02] ☐ (b) Electricity

[03] ☐ (c) Natural gas from underground pipes

[04] ☐ (d) Fuel oil

[05] ☐ (e) Propane, LPG (Bottled or tank gas)

[06] ☐ (f) Kerosene

[08] ☐ (g) Solar collectors

[10] ☐ (h) Heat pump

[20] ☐ (i) Other (Describe: _____)

[98] ☐ Don't Know

[97] ☐ Not Applicable

237

86. How many shower facilities, if any, are there in your home?
Include all showers, even those not used.

259
VAR

_____ Number of showers
[0] ☐ None—SKIP TO Q-88 238

IF "ONE OR MORE SHOWERS," ASK:

87. How many of these showers, if any, have a flow restrictor on the outlet,
either low-flow showerheads or low-flow inserts?

_____ Number, low-flow showerheads 239
_____ Number, low-flow inserts 240

[8] ☐ Don't Know

INTERVIEWER INSTRUCTIONS:
If "manually adjustable" showerheads,
probe for usage.

88. How many of the sink faucets in your home have low-flow devices installed?

_____ Number of low-flow devices 241
[00] ☐ None

89. Have you ever permanently removed any low-flow shower or faucet devices in your home?

[0] ☐ No—SKIP TO Q-91
[1] ☐ Yes 242
[8] ☐ Don't Know—SKIP TO Q-91

IF "YES," ASK:

90. What was your main reason for removing the device(s)?

[1] ☐ Not enough flow
[2] ☐ Uneven flow
[3] ☐ Clogged up
[4] ☐ Did not fit
[5] ☐ Other (SPECIFY): _____ 243

[7] ☐ Not Applicable

91. Do you have air conditioning in your home?

[0] ☐ No—SKIP TO Q-98

[1] ☐ Yes

244

IF "YES," ASK:

92. From the list on Exhibit #92, please indicate what type of air conditioning system you have in your home.

[1] ☐ (a) Central electric system

[2] ☐ (b) Central gas system

[3] ☐ (c) Evaporative swamp cooler
—SKIP TO Q-95

[4] ☐ (d) Heat pump

[5] ☐ (e) Individual window or wall units

[7] ☐ Not Applicable

245

IF "INDIVIDUAL WINDOW OR WALL UNITS," ASK:

93. How many window or wall units do you have in your home?

_____ Number of units

SKIP TO Q-95

[7] ☐ Not Applicable

246

IF 2 OR MORE HOUSING UNITS IN BUILDING, ASK:

94. Is it a central air conditioning system for your building (or group of buildings), or is it the main air conditioning equipment for your living quarters only?

[1] ☐ Central system for building

[2] ☐ Air conditioning is for
these living quarters only

[8] ☐ Don't Know

[7] ☐ Not Applicable

247

VAR

INTERVIEWER INSTRUCTIONS:

For each of the following questions, treat "Central System" as "AC#1." Treat "Window or Wall Unit" which is used most as "AC#1" and, if present, the second most used as "AC#2."

95. Which of the statements on Exhibit #95 best describes the way you used your air conditioner(s) during the months when cooling may be required?

- | | AC#1 | AC#2 | |
|---|--------------------------|--------------------------|-----|
| (a) Did not use at all | <input type="checkbox"/> | <input type="checkbox"/> | [0] |
| (b) Turned on only a few days or nights when really needed | <input type="checkbox"/> | <input type="checkbox"/> | [1] |
| (c) Turned on a moderate amount of time during the cooling season | <input type="checkbox"/> | <input type="checkbox"/> | [2] |
| (d) Turned on just about all of the cooling season | <input type="checkbox"/> | <input type="checkbox"/> | [3] |
| (e) Other (Describe: _____) . | <input type="checkbox"/> | <input type="checkbox"/> | [4] |
| Not Applicable | <input type="checkbox"/> | <input type="checkbox"/> | [7] |

96. Was this air conditioning equipment already installed in your home when you (member[s] of your household) first moved in?

- | | | | |
|-------------------|--------------------------|--------------------------|-----|
| No | <input type="checkbox"/> | <input type="checkbox"/> | [0] |
| Yes | <input type="checkbox"/> | <input type="checkbox"/> | [1] |
| Not Applicable .. | <input type="checkbox"/> | <input type="checkbox"/> | [7] |

97. Please turn to Exhibit #97 and indicate the category which best describes how many years old this air conditioning equipment is.

- | | | | |
|----------------------------|--------------------------|--------------------------|-----|
| (a) Less than 1 year .. | <input type="checkbox"/> | <input type="checkbox"/> | [1] |
| (b) 1 to 2 years | <input type="checkbox"/> | <input type="checkbox"/> | [2] |
| (c) 3 to 5 years | <input type="checkbox"/> | <input type="checkbox"/> | [3] |
| (d) 6 to 10 years | <input type="checkbox"/> | <input type="checkbox"/> | [4] |
| (e) More than 10 years ... | <input type="checkbox"/> | <input type="checkbox"/> | [5] |
| Don't Know | <input type="checkbox"/> | <input type="checkbox"/> | [8] |
| Not Applicable | <input type="checkbox"/> | <input type="checkbox"/> | [7] |

ASK EVERYONE:

VAR

Now I would like to ask you some questions about the appliances you have here in your home.

The first few questions are about your televisions.

98. How many televisions do you have that you use here in your home, at least occasionally?

_____ Number of televisions

[00] ☐ None—SKIP TO Q-101

254

FOR EACH TELEVISION, UP TO 2,
ASK FIRST ABOUT THE TELEVISION USED MOST

		TV#1	TV#2	
99. Is this television color or black & white?	Color	255	<input type="checkbox"/>	<input type="checkbox"/> [1]
	Black & white		<input type="checkbox"/>	<input type="checkbox"/> [2]
100. Please review Exhibit #97 and indicate the category which best describes how many years old this TV is.	(a) Less than 1 year	257	<input type="checkbox"/>	<input type="checkbox"/> [1]
	(b) 1 to 2 years		<input type="checkbox"/>	<input type="checkbox"/> [2]
	(c) 3 to 5 years		<input type="checkbox"/>	<input type="checkbox"/> [3]
	(d) 6 to 10 years		<input type="checkbox"/>	<input type="checkbox"/> [4]
	(e) More than 10 years ..		<input type="checkbox"/>	<input type="checkbox"/> [5]
	Don't Know		<input type="checkbox"/>	<input type="checkbox"/> [8]

256

258

The next few questions are about the appliances you use for cooking.

VAR

101. What fuel do your stove top burners use?

- [01] ☐ Wood
- [02] ☐ Electricity
- [03] ☐ Natural gas from underground pipes
- [04] ☐ Propane, LPG (Bottled or tank gas) 259
- [06] ☐ Kerosene
- [20] ☐ Other (Describe): _____
- [00] ☐ None—no burners—SKIP TO Q-104

102. Would you please look at Exhibit #102 and indicate the category which best describes when the stove top burners were installed in your home.

- [1] ☐ (a) Already installed when I moved in
- [2] ☐ (b) Brought with me when I moved in
- [3] ☐ (c) Purchased by me when I moved in or since I moved in 260
- [4] ☐ (d) Purchased by someone else when I moved in or since I moved in
- [7] ☐ Not Applicable

103. Now would you please turn to Exhibit #103 and indicate which of the categories best describes your estimate of how many years old your stove top burners are?

- [1] ☐ (a) Less than 1 year
- [2] ☐ (b) 1 to 2 years
- [3] ☐ (c) 3 to 5 years 26
- [4] ☐ (d) 6 to 10 years
- [5] ☐ (e) More than 10 years
- [8] ☐ Don't Know
- [7] ☐ Not Applicable

104. How many ovens, including microwave or convection ovens, do you have that you use for cooking, at least occasionally? Please consider each oven separately, regardless of whether they are combined in a single unit. For example, a double-oven unit counts as two ovens. Please do not include toaster ovens.

VAR

Number of ovens
[0] ☐ None—SKIP TO Q-111

262

FOR EACH OVEN, UP TO 2,
ASK FIRST ABOUT THE OVEN USED MOST

		Oven #1	Oven #2	
105. Is this oven electric, gas, or does it use some other fuel?	Gas263	<input type="checkbox"/>	<input type="checkbox"/>	[1]
	Electric264	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	[2]
	Other	<input type="checkbox"/>	<input type="checkbox"/>	
IF "ELECTRIC," ASK:				
106. Is it a microwave oven?	No265	<input type="checkbox"/>	<input type="checkbox"/>	[0]
	Yes	<input type="checkbox"/>	<input type="checkbox"/>	[1]
	N/A	<input type="checkbox"/>	<input type="checkbox"/>	[7]
107. Is this oven separate or is it combined with your stove top burners in a single unit?	Separate267	<input type="checkbox"/>	<input type="checkbox"/>	[1]
	Combined	<input type="checkbox"/>	<input type="checkbox"/>	[2]
108. Please look at Exhibit #108 and indicate the category which best describes when the oven was installed in your home.				
(a) Already installed when I moved in269		<input type="checkbox"/>	<input type="checkbox"/>	[1]
(b) Brought with me when I moved in		<input type="checkbox"/>	<input type="checkbox"/>	[2]
(c) Purchased by me when I moved in or since I moved in		<input type="checkbox"/>	<input type="checkbox"/>	[3]
(d) Purchased by someone else when I moved in or since I moved in		<input type="checkbox"/>	<input type="checkbox"/>	[4]
109. Now would you please review Exhibit #109 and indicate the category which best describes your estimate of how many years old the oven is?	(a) Less than 1 year .271	<input type="checkbox"/>	<input type="checkbox"/>	[1]
	(b) 1 to 2 years	<input type="checkbox"/>	<input type="checkbox"/>	[2]
	(c) 3 to 5 years	<input type="checkbox"/>	<input type="checkbox"/>	[3]
	(d) 6 to 10 years	<input type="checkbox"/>	<input type="checkbox"/>	[4]
	(e) More than 10 years ..	<input type="checkbox"/>	<input type="checkbox"/>	[5]
	Don't Know	<input type="checkbox"/>	<input type="checkbox"/>	[8]
110. Is it a self-cleaning oven?	No273	<input type="checkbox"/>	<input type="checkbox"/>	[1]
	Yes	<input type="checkbox"/>	<input type="checkbox"/>	[2]

Now I would like to ask you a few questions about your refrigerators and freezers.

265

111. How many refrigerators do you have that are presently in use?

VAR

_____ Number presently in use
[0] ☐ None—SKIP TO Q-115

275

FOR EACH REFRIGERATOR, UP TO 2
ASK FIRST ABOUT THE REFRIGERATOR USED MOST

	Refrig #1	Refrig #2	
112. How large is this refrigerator?			
• Large (17 cubic feet or more) 276	<input type="checkbox"/>	<input type="checkbox"/>	[1] 277
• Medium (10 to 16 cubic feet)	<input type="checkbox"/>	<input type="checkbox"/>	[2]
• Small (less than 10 cubic feet) ..	<input type="checkbox"/>	<input type="checkbox"/>	[3]
113. Which of the following best describes the location of the freezer section?			
• Side-by-side 278	<input type="checkbox"/>	<input type="checkbox"/>	[1] 279
• Top	<input type="checkbox"/>	<input type="checkbox"/>	[2]
• Bottom	<input type="checkbox"/>	<input type="checkbox"/>	[3]
• Inside	<input type="checkbox"/>	<input type="checkbox"/>	[4]
• No freezer section ...	<input type="checkbox"/>	<input type="checkbox"/>	[5]
114. Is this refrigerator "frost-free" or must it be manually defrosted?			
• Frost-free 280	<input type="checkbox"/>	<input type="checkbox"/>	[1] 281
• Manual defrost	<input type="checkbox"/>	<input type="checkbox"/>	[2]
• No freezer section ...	<input type="checkbox"/>	<input type="checkbox"/>	[5]

115. How many freezers which are separate from your refrigerator do you have that are presently in use?

282

_____ Number in use
[0] ☐ None—SKIP TO Q-119

FOR EACH SEPARATE FREEZER, UP TO 2,
ASK FIRST ABOUT THE FREEZER USED MOST:

	Freezer #1	Freezer #2	
116. How large is this freezer?			
• Large (17 cubic feet or more) 283	<input type="checkbox"/>	<input type="checkbox"/>	[1] 284
• Medium (10 to 16 cubic feet)	<input type="checkbox"/>	<input type="checkbox"/>	[2]
• Small (less than 10 cubic feet) ..	<input type="checkbox"/>	<input type="checkbox"/>	[3]
117. Is this an upright freezer or a horizontal (chest-type) freezer?			
• Upright 285	<input type="checkbox"/>	<input type="checkbox"/>	[1] 286
• Horizontal (chest)	<input type="checkbox"/>	<input type="checkbox"/>	[2]
118. Is the freezer "frost-free" or is it "manual defrost"?			
• Frost-free 287	<input type="checkbox"/>	<input type="checkbox"/>	[1] 288
• Manual defrost	<input type="checkbox"/>	<input type="checkbox"/>	[2]

VAR

Next, I would like to know about some of the major home appliances which you use here in your home.

119. Would you please review the list of appliances on Exhibit #119. For each of the appliances listed, please indicate whether you use the appliance on your own home electric or gas meter lines.

- | No
[0] | Yes
[1] | | |
|--------------------------|-------------------------------------|-----------------------------|-----|
| <input type="checkbox"/> | <input type="checkbox"/> | (a) Clothes washing machine | 289 |
| <input type="checkbox"/> | <input type="checkbox"/> | (b) Electric clothes dryer | 290 |
| <input type="checkbox"/> | <input type="checkbox"/> | (c) Gas clothes dryer | 291 |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | (d) Electric dishwasher | 292 |

IF "DISHWASHER" IS USED, ASK:

120. Does your electric dishwasher have either of the following?

- | No
[0] | Yes
[1] | OK
[8] | | |
|--------------------------|-------------------------------------|--------------------------|----------------------------------|-----|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Its own water heating element | 293 |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | A separate "energy saver switch" | 294 |

IF "YES," ASK:

121. Do you use the switch regularly?

- | | | | |
|-----|--------------------------|------------|-----|
| [0] | <input type="checkbox"/> | No | |
| [1] | <input type="checkbox"/> | Yes | 295 |
| [8] | <input type="checkbox"/> | Don't Know | |

122. Is the heating element for the drying cycle disconnected or do you regularly stop the dishwasher at the beginning of the drying cycle?

- | | | | |
|-----|--------------------------|---------------------------------|-----|
| [1] | <input type="checkbox"/> | Dry cycle disconnected | |
| [2] | <input type="checkbox"/> | Regularly stop before dry cycle | 296 |
| [0] | <input type="checkbox"/> | Neither | |
| [8] | <input type="checkbox"/> | Don't Know | |

IF "2 OR MORE UNITS," SKIP TO Q-125

VAR

IF MOBILE HOME OR SINGLE FAMILY DWELLING, ASK:

123. Do you have your own swimming pool, hot tub or Jacuzzi?
MARK ALL THAT APPLY.

	No	Yes
	[0]	[1]
Swimming pool	<input type="checkbox"/>	<input type="checkbox"/>
Hot tub	<input type="checkbox"/>	<input type="checkbox"/>
Jacuzzi	<input type="checkbox"/>	<input type="checkbox"/>

297

298

299

FOR EACH "YES," ASK:

124. Please turn to Exhibit #124 and indicate which fuel is used to heat the (pool/hot tub/Jacuzzi).

	300	301	302	
	Pool	Hot Tub	Jacuzzi	
(a) None—not heated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[00]
(b) Electricity	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[02]
(c) Natural gas from underground pipes ...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[03]
(d) Fuel oil	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[04]
(e) Propane, LPG (Bottled or tank gas)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[05]
(f) Solar collectors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[08]
(g) Other (Describe: _____)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[20]
Not Applicable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[97]

INTERVIEWER INSTRUCTIONS:

A "heated" pool, hot tub or Jacuzzi, is one which uses some sort of mechanical heating equipment. Passive solar, including "solar blanket" does not count as a "heated" pool/hot tub/Jacuzzi.

ASK EVERYONE:

125. How many waterbed heaters do you have which are presently in use?

_____ Number presently in use 303

126. Would you please review the list of equipment on Exhibit #126 and indicate whether you regularly use any of these types of equipment on your household electric or gas meter lines.

No [0]	Yes [1]		
<input type="checkbox"/>	<input type="checkbox"/>	(a) Special large health or medical equipment	304
<input type="checkbox"/>	<input type="checkbox"/>	(b) Photography laboratory equipment	305
<input type="checkbox"/>	<input type="checkbox"/>	(c) Woodworking equipment (including saw, planers, etc.)	306
<input type="checkbox"/>	<input type="checkbox"/>	(d) Electric ceramic kiln	307
<input type="checkbox"/>	<input type="checkbox"/>	(e) Gas ceramic kiln	308
<input type="checkbox"/>	<input type="checkbox"/>	(f) Home computer or video game	309
<input type="checkbox"/>	<input type="checkbox"/>	(g) Electric office equipment	310
<input type="checkbox"/>	<input type="checkbox"/>	(h) Metalworking equipment (including welding)	311
<input type="checkbox"/>	<input type="checkbox"/>	(i) Lighting or heating for greenhouse	312
<input type="checkbox"/>	<input type="checkbox"/>	(j) Water pumps for irrigation	313
<input type="checkbox"/>	<input type="checkbox"/>	(k) Electric well water pump for household water	314

127. Do any of your household electric or gas bills include charges for fuel used for purposes other than your own living quarters, such as farm buildings, or machinery, the house or apartment of another household, a business or office, or any other large uses?

[0] ☐ No 315
 [1] ☐ Yes

IF "YES," ASK:

128. Which of your fuel bills include charges for fuel used for purposes other than your own living quarters?

No [0]	Yes [1]		
<input type="checkbox"/>	<input type="checkbox"/>	(a) Electricity	316
<input type="checkbox"/>	<input type="checkbox"/>	(b) Natural gas from underground pipes	317

In order to help us better understand your answers to the questions about your home and how you use energy, and to help the Bonneville Power Administration to better design programs to fit the needs of different people, I have a few questions about yourself and the other people who live here.

129. First, how many persons, including yourself, normally reside here—
that is, those who live here for more than half of the year?

_____ Number of residents

318

130. Now, would you please tell me who the people are who live here—just in relation to yourself. I would also like to know the ages of each of the people.
If you prefer, you may use the list of age categories on Exhibit #130.

Person Number	Relationship to Respondent	Code	Gender		Age	
			M [0]	F [1]	Actual	Category
1	PRIMARY RESPONDENT	319 1	320 <input type="checkbox"/>	<input type="checkbox"/>	321	322
2		323	324 <input type="checkbox"/>	<input type="checkbox"/>	325	326
3		327	328 <input type="checkbox"/>	<input type="checkbox"/>	329	330
4		331	332 <input type="checkbox"/>	<input type="checkbox"/>	333	334
5		335	336 <input type="checkbox"/>	<input type="checkbox"/>	337	338
6		339	340 <input type="checkbox"/>	<input type="checkbox"/>	341	342
7		343	344 <input type="checkbox"/>	<input type="checkbox"/>	345	346
8		347	348 <input type="checkbox"/>	<input type="checkbox"/>	349	350
9		351	352 <input type="checkbox"/>	<input type="checkbox"/>	353	354
10		355	356 <input type="checkbox"/>	<input type="checkbox"/>	357	358

INTERVIEWER INSTRUCTIONS:

Relationship Code—

- Primary respondent —1
- Spouse of respondent —2
- Parent of respondent —3
- Child of respondent —4
- Other relative —5
- Other non-relative —6

131. RECORD: Did HH#2 actively participate in the interview?

VAR

- [0] ☐ No, not present
 [1] ☐ No, present but did not participate 359
 [2] ☐ Yes, part time
 [3] ☐ Yes, full time

132. Please review Exhibit #132 and indicate which category of race and ethnic origin best describes (HOUSEHOLDER #1). What about (HOUSEHOLDER #2)?

- | | HH#1 | HH#2 | |
|-----|-------------------------------|--|-----|
| 360 | <input type="checkbox"/> [01] | <input type="checkbox"/> (a) American Indian or Alaskan Native | 361 |
| | <input type="checkbox"/> [02] | <input type="checkbox"/> (b) Asian or Pacific Islander | |
| | <input type="checkbox"/> [03] | <input type="checkbox"/> (c) Black, not of Hispanic Origin | |
| | <input type="checkbox"/> [04] | <input type="checkbox"/> (d) Black Hispanic | |
| | <input type="checkbox"/> [05] | <input type="checkbox"/> (e) White Hispanic | |
| | <input type="checkbox"/> [06] | <input type="checkbox"/> (f) White, not of Hispanic Origin | |
| | <input type="checkbox"/> [10] | <input type="checkbox"/> (g) Other: (Describe: _____) | |
| | | _____) | |
| | <input type="checkbox"/> [98] | <input type="checkbox"/> Don't Know | |
| | <input type="checkbox"/> [96] | <input type="checkbox"/> Refused | |
| | [97] | <input type="checkbox"/> Not Applicable | |

133. Please turn to Exhibit #133 and indicate the category which best describes how much formal education (HOUSEHOLDER #1) has received? Please just tell me the letter corresponding to the category which includes the last grade (or year) of school completed by (HOUSEHOLDER #1). What about (HOUSEHOLDER #2)?

- | | HH#1 | HH#2 | |
|-----|-------------------------------|--|-----|
| 362 | <input type="checkbox"/> [01] | <input type="checkbox"/> (a) Never attended school | 363 |
| | <input type="checkbox"/> [02] | <input type="checkbox"/> (b) Some elementary school | |
| | <input type="checkbox"/> [03] | <input type="checkbox"/> (c) Completed elementary school | |
| | <input type="checkbox"/> [04] | <input type="checkbox"/> (d) Some high school | |
| | <input type="checkbox"/> [05] | <input type="checkbox"/> (e) Completed high school | |
| | <input type="checkbox"/> [06] | <input type="checkbox"/> (f) Completed trade/vocational school | |
| | <input type="checkbox"/> [07] | <input type="checkbox"/> (g) Some college | |
| | <input type="checkbox"/> [08] | <input type="checkbox"/> (h) Completed college | |
| | <input type="checkbox"/> [09] | <input type="checkbox"/> (i) Some post-graduate | |
| | <input type="checkbox"/> [10] | <input type="checkbox"/> (j) Completed post-graduate | |
| | <input type="checkbox"/> [98] | <input type="checkbox"/> Don't Know | |
| | <input type="checkbox"/> [99] | <input type="checkbox"/> Refused | |
| | [97] | <input type="checkbox"/> Not Applicable | |

134. Now please turn to Exhibit #134 and indicate the letter of the category which best describes the total combined income before taxes for your household for 1982.

VAR

EXHIBIT #134

- (a) UNDER \$5,000
- (b) \$ 5,000 - \$ 7,999
- (c) \$ 8,000 - \$ 9,999
- (d) \$10,000 - \$11,999
- (e) \$12,000 - \$13,999
- (f) \$14,000 - \$15,999
- (g) \$16,000 - \$17,999
- (h) \$18,000 - \$19,999
- (i) \$20,000 - \$24,999
- (j) \$25,000 - \$29,999
- (k) \$30,000 - \$34,999
- (l) \$35,000 - \$39,999
- (m) \$40,000 - \$49,999
- (n) \$50,000 - \$59,999
- (o) \$60,000 - \$74,999
- (p) \$75,000 OR OVER

- (a) ☐ [01]
- (b) ☐ [02]
- (c) ☐ [03]
- (d) ☐ [04]
- (e) ☐ [05]
- (f) ☐ [06]
- (g) ☐ [07]
- (h) ☐ [08]
- (i) ☐ [09]
- (j) ☐ [10]
- (k) ☐ [11]
- (l) ☐ [12]
- (m) ☐ [13]
- (n) ☐ [14]
- (o) ☐ [15]
- (p) ☐ [16]

Don't Know ☐ [98]Refused ☐ [96]

364

VAR

135. We may have covered some of these points before, but just to be sure, please look at Exhibit #135 and indicate whether these fuels are used for these purposes in your household.

	Not Used [0]	Used [1]	Paid By Household [1]	Included in Rent [2]	Other [3]	
<u>ELECTRICITY</u>						
(a) For hot water	365	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	366
(b) For heating your home	367	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	368
(c) For air-conditioning (central or window/wall units)	369	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	370
(d) For cooking	371	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	372
(e) For lighting and other appliances	373	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	374
<u>GAS FROM UNDERGROUND PIPES SERVING YOUR NEIGHBORHOOD</u>						
(f) For hot water	375	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	376
(g) For heating your home	377	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	378
(h) For air-conditioning (central or window/wall units)	379	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	380
(i) For cooking	381	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	382
(j) For other appliances (including outdoor lighting)	383	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	384
<u>LP GAS (BOTTLED OR TANK GAS)</u>						
(k) For hot water	385	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	386
(l) For heating your home	387	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	388
(m) For air-conditioning (central or window/wall units)	389	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	390
(n) For cooking	391	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	392
(o) For other appliances	393	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	394
<u>FUEL OIL OR KERSOENE</u>						
(p) For hot water	395	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	396
(q) For heating your home	397	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	398

FOR EACH USE OF EACH FUEL, ASK:

136. Is that paid for by your household, included in your rent, or do you get it some other way?

273

IF UNDERGROUND GAS IS NOT USED, ASK:

VAR

137. Is gas from underground pipes available in this neighborhood?

- [0] ☐ No
[1] ☐ Yes
[8] ☐ Don't Know

399

IF ANY FUEL BILLS ARE PAID BY HOUSEHOLD, ASK:

138. Is your electricity billed on a "budget" basis? That is, is your electricity bill the same each month?

- [0] ☐ No (varies)
[1] ☐ Yes (same)
[7] ☐ Not Applicable

400

IF "NATURAL GAS" USED, ASK:

139. Is your natural gas billed on a "budget" basis?

- [0] ☐ No (varies)
[1] ☐ Yes (same)
[7] ☐ Not Applicable

401

PACIFIC NORTHWEST RESIDENTIAL ENERGY CONSUMPTION SURVEY

AUTHORIZATION FORM FOR UTILITY CONSUMPTION INFORMATION

Department of Energy
Bonneville Power Administration
P.O. Box 3621
Portland, Oregon 97208

I hereby give permission to the company or companies listed below to provide information to Louis Harris and Associates for use in connection with their survey for the Bonneville Power Administration.

This authorization covers the quantity and price of electricity and natural gas purchased by my household beginning September 1981 and continuing through December 1982. Companies are authorized to provide this information by monthly or bimonthly periods, whichever applies.

A photocopy of this authorization may be accepted with the same authority as the original.

I understand that this energy quantity and price information will be linked with my answers to the interview questions. I also understand that the energy information and interview answers may be provided to my local electric utility or natural gas company (where applicable), and that these companies may be able to identify me. However, this information shall be used for statistical purposes only. My name will never be published as a participant in this survey and I will never be contacted for advertising or promotional purposes. The electric utility or natural gas company agrees to limit access to this information and to treat me no differently than they do all of their residential customers.

I understand that I may be contacted in future surveys conducted for the Bonneville Power Administration. My participation in those future surveys will be completely voluntary and independent of my participation in this survey.

I have read and understood the above statements and agree to their provisions.

Date: _____ Signature: _____

PLEASE
PRINT

BILLING NAME _____	
SERVICE ADDRESS _____	APT. NO. _____
CITY _____	STATE _____ ZIP _____
TELEPHONE (AREA CODE) _____ NUMBER _____	

PLEASE COMPLETE ONE BLOCK BELOW FOR EACH FUEL USED BY YOUR HOUSEHOLD.

ELECTRICITY

PRINT FULL NAME OF ELECTRIC COMPANY	ACCOUNT NUMBER
-------------------------------------	----------------

NATURAL GAS
FROM UNDER-
GROUND PIPES

PRINT FULL NAME OF GAS COMPANY	ACCOUNT NUMBER
--------------------------------	----------------

VAR

TIME: _____	A.M. P.M.
-------------	--------------

0-1. In order to complete this interview, there are two other tasks that I would like to perform. First, would you mind if I measure the temperature of the "hot" tap water in your home? I can do this at any sink which has a "hot" water tap.

RECORD: Temperature of hot water _____ °F

402

[997] ☐ No hot water

[996] ☐ Refused

0-2. Has any hot water been used in the last hour for things like clothes washing, dishwashing, bathing or other tasks which use large amounts of hot water?

[0] ☐ No

403

[1] ☐ Yes

IF "YES," ASK:

0-3. What was it used for?

	No [0]	Yes [1]
a. Clothes washing	<input type="checkbox"/>	<input type="checkbox"/>
b. Dish washing	<input type="checkbox"/>	<input type="checkbox"/>
c. Bathing	<input type="checkbox"/>	<input type="checkbox"/>
d. Other	<input type="checkbox"/>	<input type="checkbox"/>

404

405

406

407

0-4. Finally, one of the most important things affecting people's energy use is the size of their home. With your permission, I would like to measure your home. Your assistance in this task would be appreciated.

In order to accurately measure your home, I would like to know about a few of the features of your home.

0-5. First, how many floors are used as living quarters?

[01] ☐ One floor

[06] ☐ More than 3 floors

[02] ☐ 1½ floors

[11] ☐ Split level—2 levels

[03] ☐ 2 floors

[12] ☐ Split level—3 levels

[04] ☐ 2½ floors

[13] ☐ Split level—More than 3 levels

[05] ☐ 3 floors

[21] ☐ Other—Describe: _____

408

0-6. How many floors that are used as living quarters are completely below ground?

VAR

- [0] ☐ None [2] ☐ One floor
 [1] ☐ Partial floor [3] ☐ More than one floor

409

0-7. How many floors that are used as living quarters are partially below ground—for example, a daylight basement?

- [0] ☐ None [2] ☐ One floor
 [1] ☐ Partial floor [3] ☐ More than one floor

410

IF "BUILDING WITH 2 OR MORE UNITS," ASK:

0-8 On what floor is this unit located, starting with the first floor which contains dwelling units?

411

Floor Number _____

0-9. What type of material is primarily used on the outside walls of this house (building)?

IF TWO MATERIALS ARE USED ABOUT THE SAME AMOUNT, RECORD BOTH MATERIALS.

Not
Used
[0]

Used
[1]

- | | | |
|---|--------------------------------------|-----|
| <input type="checkbox"/> <input type="checkbox"/> | Wood | 412 |
| <input type="checkbox"/> <input type="checkbox"/> | Brick | 413 |
| <input type="checkbox"/> <input type="checkbox"/> | Stone | 414 |
| <input type="checkbox"/> <input type="checkbox"/> | Concrete | 415 |
| <input type="checkbox"/> <input type="checkbox"/> | Stucco | 416 |
| <input type="checkbox"/> <input type="checkbox"/> | Aluminum Siding | 417 |
| <input type="checkbox"/> <input type="checkbox"/> | Steel Siding | 418 |
| <input type="checkbox"/> <input type="checkbox"/> | Composition Siding (Asbestos, Vinyl) | 419 |
| <input type="checkbox"/> <input type="checkbox"/> | Glass | 420 |
| <input type="checkbox"/> <input type="checkbox"/> | Other (Describe: _____) | 421 |

INTERVIEWER INSTRUCTIONS:

In general, measure all parts of the housing unit which are used as year-round living quarters.

- **Basements**

Include basements in one-family houses if at least some portion is heated (refer to Q-37).

Include basement space in buildings with 2 to 4 housing units, if at least some portion is heated and it is for the exclusive use of the household interviewed.

Exclude basements in one-family houses and buildings with 2 to 4 units if no portion of the basement is heated.

Exclude basements in buildings with 5 or more units.

Exclude crawl spaces.

- **Attics**

Include attics if heated or finished. Check appropriate box on corresponding diagram (refer to Q-17/18).

Exclude attics if unheated and also unfinished.

- **Garages, sheds or barns**

Include garages only if attached to house, enclosed from the weather and heated.

Exclude garages, sheds, or barns if not attached to house or open to the weather or unheated.

- **Porches**

Include porches if enclosed from the weather and heated.

Exclude porches if open to the weather or unheated.

- **Buildings with 2 or more housing units**

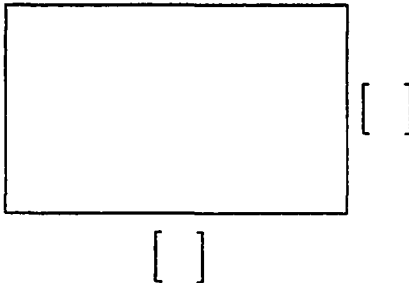
Measure only the space used exclusively by the interviewed household. Do not measure the entire building. Do not measure entryways, hallways, or any other areas which are not for the exclusive use of the interviewed household.

USE BACKS OF MEASUREMENT PAGES FOR ADDITIONAL SPACE AS NEEDED, FOR SKETCHES AND MEASUREMENTS.

RECORD MEASUREMENTS ON DIAGRAM TO NEAREST FOOT

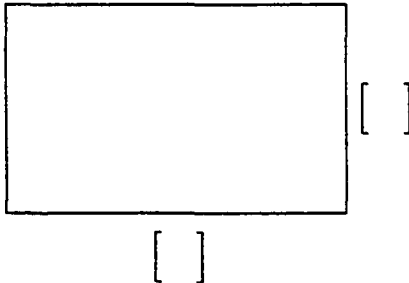
START HERE

if this household has
a basement or cellar
(see instruction on
facing page for
basements and
cellars)

Basement Measurements [1] <input type="checkbox"/> Full Basement [2] <input type="checkbox"/> Half Basement	
RECTANGULAR SHAPE 	DRAW DIAGRAM, IF OTHER THAN RECTANGULAR

START HERE

if this household
does not have a
basement or cellar

First Story Measurements [1] <input type="checkbox"/> Full Story [2] <input type="checkbox"/> Half Story	
RECTANGULAR SHAPE 	DRAW DIAGRAM, IF OTHER THAN RECTANGULAR

CONTINUE ON PAGE 51
FOR SECOND AND
THIRD STORIES

FOR OFFICE USE ONLY

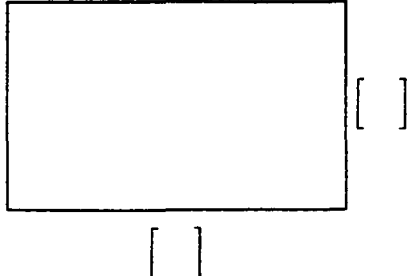
B

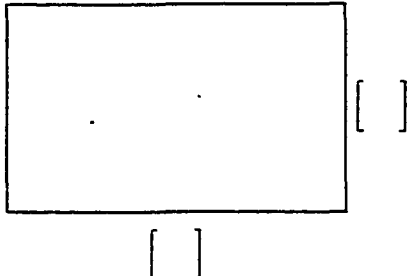
1

Full-Half	Unit A		Unit B		Unit C		Unit D		Unit X	# of Units
422	423	424	425	426	427	428	429	430	431	432
433	434	435	436	437	438	439	440	441	442	443

IF NO SECOND OR THIRD STORY TO MEASURE, GO TO Q-R-1

RECORD MEASUREMENTS ON DIAGRAM TO NEAREST FOOT

Second Story Measurements [1] <input type="checkbox"/> Full Story [2] <input type="checkbox"/> Half Story	Attic [0] <input type="checkbox"/> No [1] <input type="checkbox"/> Yes
RECTANGULAR SHAPE	DRAW DIAGRAM, IF OTHER THAN RECTANGULAR
	

Third Story Measurements [1] <input type="checkbox"/> Full Story [2] <input type="checkbox"/> Half Story	Attic [0] <input type="checkbox"/> No [1] <input type="checkbox"/> Yes
RECTANGULAR SHAPE	DRAW DIAGRAM, IF OTHER THAN RECTANGULAR
	

FOR OFFICE USE ONLY

2

3

Full-Half	Attic	Unit A		Unit B		Unit C		Unit D		Unit X	# of Units
444	445	446	447	448	449	450	451	452	453	454	455
456	457	458	459	460	461	462	463	464	465	466	467

TOTALS

Basement	First	Second	Third	Misc.	Grand
468	469	470	471	472	473

VAR

INTERVIEWER REPORT ON MEASUREMENT OF DWELLING UNIT:**R-1. HOW WERE MEASUREMENTS OBTAINED?**

- [1] ☐ Measured outside
- [2] ☐ Measured inside
- [3] ☐ Combination of outside and inside measurements
- [4] ☐ Respondent gave total square feet from house plans
- [5] ☐ Respondent estimated square feet
- [6] ☐ Other procedure (Describe: _____)
- [9] ☐ Refused

474

IF "RESPONDENT" SOURCE:

R-2. RECORD: Total square feet _____

R-3. WHAT PROBLEMS, IF ANY, DID YOU HAVE IN MEASURING THIS DWELLING UNIT?**R-4. WHAT EFFECT, IF ANY, DID THESE PROBLEMS HAVE ON THE ACCURACY OF YOUR MEASUREMENTS?**

281

DATE:

VAR

Day _____

475

Month _____

476

TIME:

Begin _____

477

AM ☐ [1]

478

PM ☐ [2]

End Interview/
Begin Measure _____

479

AM ☐ [1]

480

PM ☐ [2]

End Measure _____

481

AM ☐ [1]

482

PM ☐ [2]

INTERVIEWER:

ID #

483

Signature _____

BILLING HISTORY DATA

The billing history data file is divided into two parts. The first part of the data file contains the electricity billing histories data, and the natural gas billing data is contained in the second part of the file.

Each part of the file contains the basic data base and case identification information, and is directly comparable to the identification information contained in the interview data file.

The format of the two billing history portions of the file are directly comparable to one another. To assist the user, the labeling and numbering schemes for the variables are equivalent except for a preceding "E" (electricity) or "G" (natural gas). Thus "EVAR701" denotes the "Beginning Date" for electricity billing period 1, while "GVAR701" denotes the "Beginning Date" for gas billing period 1.

To conserve space, a generic list of the variables is included here; the preceding "E" or "G" has been omitted from the Referant or "VAR Name". Please note that in using the data, it is necessary to include these prefixes.

VARIABLE LIST

<u>Referant (VAR Name)</u>	<u>Definition</u>
ID	Serial case identification number
SOURCE	
DBNUM	
EDITION	Data Base Number
DBTYPE	
DATE	
SAMPTYP	
UTIL	
GEO	Sample Point Number
MRR	
SITE	
UBILL	Fuel and Unit Identification
	1 = Electricity, KWh
	2 = Gas, Therms
	3 = Gas, BTU
BILLPD	Billing Period
	1 = Monthly
	2 = Bimonthly
	3 = Tri-monthly
	4 = Quarterly
	5 = Semi-annually
	6 = Annually

<u>Referant (VAR Name)</u>	<u>Definition</u>
VAR701 to VAR720	Beginning Date, Billing Period 1 to Period 20
VAR721 to VAR740	Consumption, Billing Period 1 to Period 20
VAR741 to VAR760	Cost, Billing Period 1 to Period 20
VAR761	End Date, Last Valid Billing Period
VAR762	Whether Data Provided <ul style="list-style-type: none"> 1 = Data Present 2 = Invalid or no waiver 3 = Data not available 4 = Other
VAR763	Whether Billed on a Budget Basis <ul style="list-style-type: none"> 1 = Budget Basis 2 = Billed as Budget, Actual Readings Recorded 3 = Actual Basis
VAR764	Whether Utility has Provided an Energy Audit <ul style="list-style-type: none"> 1 = Audit Provided 2 = Audit Requested, not yet Provided 3 = No Data Available in Utility Records 4 = No Audit Provided or Requested
VAR765	Type of Assistance <ul style="list-style-type: none"> 1 = Zero Interest Loan From Utility 2 = BPA Buyback Program Participant 3 = Water Heater Wrap 4 = Grant 5 = BPA Weatherization Program Participant 6 = Other Weatherization Assistance 7 = No Assistance Provided 8 = Other
VAR766	Servicing Gas Company

NOAA WEATHER DATA

Heating and cooling degree days have been calculated using National Oceanographic and Atmospheric Administration (NOAA) data for climatological divisions (CD). The degree days have been calculated separately for electricity billing periods and natural gas billing periods, and are divided into two distinct computer files.

For each fuel, heating degree day (HDD) and cooling degree day (CDD) data are provided for each billing period for which consumption data are available. The applicable time frame (September 1981 through May 1983) has been divided into 20 such periods.

The data for each case are divided into four parts, each corresponding to a pair of HDD/CDD Bases, as follows:

	Heating Degree Days	Cooling Degree Days
1.	55	65
2.	60	70
3.	65	75
4.	70	80

Each record contains the serial identification code, the CD code, and the relevant HDD and CDD base temperatures, in addition to the actual HDD and CDD data for each applicable fuel billing period.

The format of the two files is identical, as is the format of each of the four records per case. As with the billing history data definitions, the labels for the electricity billing periods weather data employ a preceding "E" while the labels for the natural gas weather data employ a preceding "G".

<u>Referant (VAR Name)</u>	<u>Definition</u>
ID	Serial case identification number
CD	NOAA Climatological Division Code
HDD1B	Heating Degree Day Base, #1 (55 degree F)
CDD1B	Cooling Degree Day Base, #1 (65 degree F)
HDD1001 - HDD1020	HDD Data for Base #1 for all Billing Periods
CDD1001 - CDD1020	CDD Data for Base #1 for all Billing Periods

This information is then repeated for each of the four degree day bases.

APPENDIX IV

FILE LABEL	PACIFIC NORTHWEST 1983 RESIDENTIAL SURVEY DATA
TITLE	EXAMPLE DATA DEFINITION FILE BASED ON PNRES83
DATA SET	
DATA LIST	FILE = RAWDATA, RECORDS = 13 /
VARIABLE LABELS	
	CD1 'FIRST RECORD IDENTIFICATION'
	ID1 'CASE IDENTIFICATION ON RECORD ONE'
	SOURCE 'ORGANIZATION SPONSORING DATA
COLLECTION'	
	DBNUMBER 'DATA BASE NUMBER'
	EDITION 'DATA BASE EDITION'
	DBTYPE 'TYPE OF DATA'
	DATE 'DATE SURVEY BEGAN'
	SAMPTYPE 'TYPE OF SAMPLE'
	UTIL 'UTILITY ID'
	EUTILITY 'ELECTRIC UTILITY RECODE OF UTIL -
MOVE?'	
	GEOAREA 'GEOGRAPHIC AREA'
	MRR 'UTILITY METER READING ROUTE'
	SITE 'SITE ID'
	VAR003 'Q1A-SHARE ELECTRIC METER'
	VAR004 'Q1B-SEPARATE COOKING FACILITIES'
	VAR005 'Q2-TYPE OF DWELLING'
	VAR006 'Q3-NUMBER OF HOUSING UNITS IN
BUILDING'	
	VAR007 'Q4-DETACHED OR ATTACHED IF 2-4
UNITS'	
	VAR008 'Q5-MEANS OF PAYMENT FOR HOUSING'
	VAR009 'Q6-FURNISHED WITH REFRIGERATOR'
	VAR010 'Q6-FURNISHED WITH BURNERS'
	VAR011 'Q6-FURNISHED WITH OVEN'
	VAR012 'Q6-FURNISHED WITH CLOTHES WASHER'
	VAR013 'Q6-FURNISHED WITH CLOTHES DRYER'
	VAR014 'Q6-FURNISHED WITH DISHWASHER'
	VAR015 'Q7-BEGINNING YEAR OF PRESENT
RESIDENCE'	
	VAR016 'Q8-EST START YEAR OF PRESENT
RESIDENCE'	
	VAR017 'Q9-HHLDERS PRIOR TO MOVE-IN'
	VAR018 'Q10-START YEAR FOR RESIDENCE OF
OTHERS'	

OTHERS '	VAR019	'Q11-EST YEAR FOR RESIDENCE OF
	VAR020	'Q12-IF 1981-1983 THEN YEAR'
RESIDENCE '	VAR021	'Q12-BEGINNING MONTH OF PRESENT
	VAR022	'Q13-NEW HOUSE WHEN FIRST OCCUPIED'
	VAR023	'Q14-YEAR DWELLING BUILT'
	VAR024	'Q15-YEAR DWELLING BUILT-ESTIM'
RENOVATED '	VAR025	'Q16-DWELLING SUBSTANTIALLY
	CD2	'CARD NUMBER 2'
	ID2	'SERIAL CASE ID2'
	VAR026	'Q17-# OF ROOMS-LIVING ROOM'
	VAR027	'Q18-NO HEAT-# LIVING ROOMS'
	VAR028	'Q17-# OF ROOMS-DINING'
	VAR029	'Q18-NO HEAT-# DINING ROOM'
	VAR030	'Q17-# OF ROOMS-KITCHEN WITH EATING'
EATING '	VAR031	'Q18-NO HEAT-# KITCHEN WITH EATING'
	VAR032	'Q17-# OF ROOMS-KITCHEN WITHOUT
EATING '	VAR033	'Q18-NO HEAT-# KITCHEN WITHOUT
	VAR034	'Q17-# OF ROOMS-BEDROOM'
	VAR035	'Q18-NO HEAT-# BEDROOM'
	VAR036	'Q17-# OF ROOMS-DEN'
	VAR037	'Q18-NO HEAT-# DEN'
	VAR038	'Q17-# OF ROOMS-FAMILY ROOM'
	VAR039	'Q18-NO HEAT-# FAMILY ROOM'
	VAR040	'Q17-# OF ROOMS-UTILITY ROOM'
	VAR041	'Q18-NO HEAT-# UTILITY ROOM'
	VAR042	'Q17-# OF ROOMS-OTHER SMALL ROOMS'
	VAR043	'Q18-NO HEAT-# OTHER SMALL ROOMS'
	VAR044	'Q17-# OF ROOMS-OTHER LARGE ROOMS'
	VAR045	'Q18-NO HEAT-# OTHER LARGE ROOMS'
	VAR046	'Q17-NUMBER OF ROOMS IN LIVING SPACE'
ROOMS '	VAR047	'Q18-NUMBER OF CLOSED OFF-UNHEATED
	VAR048	'Q18A-ANY ROOMS NOT HEATED'
	VAR049	'Q19-ENERGY EFFICIENCY OF HOME'
	VAR050	'Q20-THOUGHTS ON ENVIRONMENTAL
POLLUTION '	VAR051	'Q20-THOUGHTS ON COST OF ENERGY'
	VAR052	'Q20-THOUGHTS ON UNEMPLOYMENT'
	VAR053	'Q20-THOUGHTS ON SCARCITY OF ENERGY'
	VAR054	'Q20-THOUGHTS ON INFLATION'
	VAR055	'Q20-THOUGHTS ON CRIME'
	VAR056	'Q21-COST PROBLEM - ELECTRICITY'
	VAR057	'Q21-SCARCITY PROBLEM - ELECTRICITY'
	VAR058	'Q21-COST PROBLEM - NATURAL GAS'
	VAR059	'Q21-SCARCITY PROBLEM - NATURAL GAS'
	VAR060	'Q21-COST PROBLEM - HEATING OIL'
	VAR061	'Q21-SCARCITY PROBLEM - HEATING OIL'

	VAR062	'Q21-COST PROBLEM - GASOLINE'
	VAR063	'Q21-SCARCITY PROBLEM - GASOLINE'
	VAR064	'Q21-COST PROBLEM - WOOD'
	VAR065	'Q21-SCARCITY PROBLEM - WOOD'
	VAR066	'Q22-STATEMENTS-COMFORTABLE AT OR <
68F'		
	VAR067	'Q22-STATEMENTS-REDUCING H2O TEMP
SAVES'		
	VAR068	'Q22-STATEMENTS-CONSERVE ENERGY=SAVES
\$'		
	VAR069	'Q22-TURNING DOWN TEMP IS WORTHWHILE'
	VAR070	'Q22-HARD TO MAKE HOME ENERGY
EFFICIENT'		
	VAR071	'Q22-RIGHT TO USE ENERGY'
	VAR072	'Q22-PRICE IS MORE IMPORTANT THEN
ENERGY'		
	VAR073	'Q22-CONSERVING ENERGY=CHANGE IN
LIFE'		
	VAR074	'Q23-HEARD OF FEDERAL TAX CREDIT'
	VAR075	'Q24-MADE USE OF FEDERAL TAX CREDIT'
	VAR076	'Q23-HEARD OF STATE TAX BENEFIT'
	VAR077	'Q24-MADE USE OF STATE TAX BENEFIT'
	VAR078	'Q23-HEARD OF LOW INTEREST LOAN'
	VAR079	'Q24-MADE USE OF LOW INTEREST LOAN'
	VAR080	'Q25-ENERGY AUDIT OF HOME'
	VAR081	'Q26-LIKELIHOOD OF AUDIT'
	VAR082	'Q27-LAST INSPECTION-MONTH'
	VAR083	'Q27-LAST INSPECTION-YEAR'
	VAR084	'Q28-ESTIMATED MOST RECENT
INSPECTION'		
	VAR085	'Q29-INSPECTION REPRESENTATIVE'
	CD3	'CARD NUMBER 3'
	ID3	'SERIAL CASE ID3'
	VAR086	'Q30A-RECOMENDATIONS-OUTSIDE WALL
INSUL'		
	VAR087	'Q30B-ACTION-OUTSIDE WALL INSUL'
	VAR088	'Q31AB-MAY 1980-OUTSIDE WALL INSUL'
	VAR089	'Q30A-RECOMENDATIONS-ROOF INSUL'
	VAR090	'Q30B-ACTION-ROOF INSUL'
	VAR091	'Q31AB-MAY 1980-ROOF INSUL'
	VAR092	'Q30A-RECOMENDATIONS-FLOOR INSUL'
	VAR093	'Q30B-ACTION-FLOOR INSUL'
	VAR094	'Q31AB-MAY 1980-FLOOR INSUL'
	VAR095	'Q30A-RECOMENDATIONS-INSUL HEAT
DUCTS'		
	VAR096	'Q30B-ACTION-INSUL HEAT DUCTS'
	VAR097	'Q31AB-MAY 1980-INSUL HEAT DUCTS'
	VAR098	'Q30A-RECOMENDATIONS-STORM DOORS'
	VAR099	'Q30B-ACTION-STORM DOORS'
	VAR100	'Q31AB-MAY 1980-STORM DOORS'
	VAR101	'Q30A-RECOMENDATIONS-STORM WINDOWS'
	VAR102	'Q30B-ACTION-STORM WINDOWS'

	VAR103	'Q31AB-MAY 1980-STORM WINDOWS'
	VAR104	'Q30A-RECOMENDATIONS-PLASTIC'
	VAR105	'Q30B-ACTION-PLASTIC'
	VAR106	'Q31AB-MAY 1980-PLASTIC'
	VAR107	'Q30A-RECOMENDATIONS-WRAP WATER HEAT'
	VAR108	'Q30B-WRAP HEATER'
	VAR109	'Q31AB-MAY 1980-WRAP WATER HEAT'
	VAR110	'Q30A-RECOMENDATIONS-INSUL WATER
PIPES'		
	VAR111	'Q30B-ACTION-INSUL WATER PIPES'
	VAR112	'Q31AB-MAY 1980-INSUL WATER PIPES'
	VAR113	'Q30A-RECOMENDATIONS-
WEATHERSTRIPPING'		
	VAR114	'Q30B-ACTION-WEATHERSTRIPPING'
	VAR115	'Q31AB-MAY 1980-WEATHERSTRIPPING'
	VAR116	'Q30A-RECOMENDATIONS-CAULKING'
	VAR117	'Q30B-ACTION-CAULKING'
	VAR118	'Q31AB-MAY 1980-CAULKING'
	VAR119	'Q30A-RECOMENDATIONS-REDUCE H2O TEMP'
	VAR120	'Q30B-ACTION-REDUCE H2O TEMP'
	VAR121	'Q31AB-MAY 1980-REDUCE H2O TEMP'
	VAR122	'Q30A-RECOMENDATIONS-REDUCE HOME
TEMP'		
	VAR123	'Q30B-ACTION-REDUCE HOME TEMP'
	VAR124	'Q31AB-MAY 1980-REDUCE HOME TEMP'
	VAR125	'Q30A-RECOMENDATIONS-AUTO SETBACK
THERM'		
	VAR126	'Q30B-ACTION-AUTO SETBACK THERM'
	VAR127	'Q31AB-MAY 1980-AUTO SETBACK THERM'
	VAR128	'Q30A-RECOMENDATIONS-OTHER'
	VAR129	'Q30B-ACTION-OTHER'
	VAR130	'Q31AB-MAY 1980-OTHER'
	VAR131	'Q32-PERCENT INSULATION-OUTSIDE
WALLS'		
	VAR132	'Q33-ADDED OUTSIDE WALL INSUL'
	VAR133	'Q34-PERCENT INSUL ROOF-CEILING-
ATTIC'		
	VAR134	'Q35-ADDED-ROOF-CEILING-ATTIC INSUL'
	VAR135	'Q36-TYPE OF FOUNDATION-BASEMENT'
	VAR136	'Q36-TYPE OF FOUNDATION-CRAWL SPACE
ENCL'		
	VAR137	'Q36-TYPE OF FOUNDATION-CRAWL SPACE
OPEN'		
	VAR138	'Q36-TYPE OF FOUNDATION-CONCRETE
SLAB'		
	VAR139	'Q36-TYPE OF FOUNDATION-OTHER'
	VAR140	'Q37-PERCENT OF BASEMENT HEATED'
	VAR141	'Q38-PERCENT AREA ABOVE BASEMENT
INSUL'		
	VAR142	'Q39-ADDED INSUL ABOVE UNHEAT
BASEMENT'		

INSUL'	VAR143	'Q40-PERCENT HEATED WALLS-BASEMENT
	VAR144	'Q41-ADDED INSULATION BASEMENT WALLS'
	CD4	'CARD NUMBER 4'
	ID4	'SERIAL CASE ID 4'
AREA'	VAR145	'Q42-NUMBER OF DOORS TO UNHEATED
	VAR146	'Q43-NUMBER OF OUTSIDE STORM DOORS'
	VAR147	'Q44-NUMBER OF SLIDING GLASS DOORS'
	VAR148	'Q45-DOORS WITH INSULATING GLASS'
	VAR149	'Q46-WINDOWS-INSULATED OR THERMOPANE'
	VAR150	'Q46-STORM WINDOWS'
	VAR151	'Q46-PLASTIC OVER WINDOWS'
	VAR152	'Q46-OTHER'
	VAR153	'Q47-PERCENT INSULATED-THERMOPANE
GLASS'	VAR154	'Q47-PERCENT STORM WINDOWS'
	VAR155	'Q47-PERCENT PLASTIC OVER WINDOWS'
	VAR156	'Q47-PERCENT OTHER'
	VAR157	'Q48-TYPE OF FUEL USED MOST'
	VAR158	'Q49-EQUIP USED MOST TO HEAT HOME-
GAS'		
	VAR159	'Q50-EQUIP USED MOST TO HEAT HOME-
ELEC'		
	VAR160	'Q51-EQUIP USED MOST TO HEAT HOME-
WOOD'		
	VAR161	'Q52-EQUIPMENT INSTALLED ON FP#1'
	VAR162	'Q52-EQUIPMENT INSTALLED ON FP#2'
	VAR163	'Q53-OWN OR SHARE HEATING SYSTEM'
	VAR164	'Q54-HEATING SYSTEM CHANGE DURING
PAST YR'		
	VAR165	'Q55-MONTH OF CHANGE'
	VAR166	'Q55-YEAR OF CHANGE'
	VAR167	'Q56-FORMER TYPE OF HEATING FUEL'
	VAR168	'Q57-HEAT EQUIP ALREADY INSTALLED'
	VAR169	'Q58-YEAR EQUIPMENT INSTALLED'
	VAR170	'Q59-ESTM YEAR HEAT EQUIP INSTALLED'
	VAR171	'Q60-USE OF OTHER HEATING FUELS'
	VAR172	'Q61-ADDITIONAL HEATING FUEL-WOOD'
	VAR173	'Q61-ADDITIONAL HEATING FUEL-ELEC'
	VAR174	'Q61-ADDITIONAL HEATING FUEL-GAS'
	VAR175	'Q61-ADDITIONAL HEATING FUEL-FUEL
OIL'		
	VAR176	'Q61-ADDITIONAL HEATING FUEL-PROPANE'
	VAR177	'Q61-ADDITIONAT HEATING FUEL-
KEROSENE'		
	VAR178	'Q61-ADDITIONAL HEATING FUEL-COAL'
	VAR179	'Q61-ADDITIONAL HEATING FUEL-SOLAR'
	VAR180	'Q61-ADDITIONAL HEATING FUEL-
GEOTHERMAL'		
	VAR181	'Q61-ADDITIONAL HEATING FUEL-OTHER'

KEROSENE'	VAR182	'Q62-TYPE OF EQUIP GAS OIL OR
	VAR183	'Q63-TYPE OF EQUIP IF ELECTRICITY'
	VAR184	'Q64-TYPE OF EQUIP IF WOOD'
	VAR185	'Q65-EQUIP FIREPLACE #1'
	VAR186	'Q65-EQUIP FIREPLACE#2'
	VAR187	'Q66-BURNED WOOD OR PRESTOLOGS PAST
YR'		
	VAR188	'Q67-NUMBER OF LOGS BURNED IN LAST
YR'		
	VAR189	'Q68-NUMBER OF CORDS BURNED IN LAST
YR'		
	VAR190	'Q69-PERCENT OF WOOD PURCHASED'
	VAR191	'Q70-THERMOSTAT ADJUSTS HEATING TEMP'
	VAR192	'Q71-TYPE OF THERMOSTAT ON HEATING
SYSTEM'		
	CD5	'CARD NUMBER 5'
	ID5	'SERIAL CASE ID 5'
	VAR193	'Q72A-TEMP OF HOUSE-SOMEONE HOME'
	VAR194	'Q72-LOW TEMP-SOMEONE HOME'
	VAR195	'Q72-HIGH TEMP-SOMEONE HOME'
	VAR196	'Q72B-TEMP OF HOUSE-NO ONE HOME'
	VAR197	'Q72-LOW TEMP-NO ONE HOME'
	VAR198	'Q72-HIGH TEMP-NO ONE HOME'
	VAR199	'Q72C-TEMP OF HOUSE-SLEEPING HOURS'
	VAR200	'Q72-LOW TEMP-SLEEPING HOURS'
	VAR201	'Q72-HIGH TEMP-SLEEPING HOURS'
	VAR202	'Q73-SOMEONE HOME 9-5 AND HOUSE
HEATED'		
1981'	VAR203	'Q74-PERIODS OF NO USE SINCE SEPT
	VAR204	'Q75-PERIOD OF NO USE-SEPT 1981'
	VAR205	'Q75-PERIOD OF NO USE-OCT 1981'
	VAR206	'Q75-PERIOD OF NO USE-NOV 1981'
	VAR207	'Q75-PERIOD OF NO USE-DEC 1981'
	VAR208	'Q75-PERIOD OF NO USE-JAN 1982'
	VAR209	'Q75-PERIOD OF NO USE-FEB 1982'
	VAR210	'Q75-PERIOD OF NO USE-MARCH 1982'
	VAR211	'Q75-PERIOD OF NO USE-APRIL 1982'
	VAR212	'Q75-PERIOD OF NO USE-MAY 1982'
	VAR213	'Q75-PERIOD OF NO USE-JUNE 1982'
	VAR214	'Q75-PERIOD OF NO USE-JULY 1982'
	VAR215	'Q75-PERIOD OF NO USE-AUG 1982'
	VAR216	'Q75-PERIOD OF NO USE-SEPT 1982'
	VAR217	'Q75-PERIOD OF NO USE-OCT 1982'
	VAR218	'Q75-PERIOD OF NO USE-NOV 1982'
	VAR219	'Q75-PERIOD OF NO USE-DEC 1982'
	VAR220	'Q75-PERIOD OF NO USE-JAN 1983'
	VAR221	'Q75-PERIOD OF NO USE-FEB 1983'
	VAR222	'Q75-PERIOD OF NO USE-MAR 1983'
	VAR223	'Q75-PERIOD OF NO USE-APRIL 1983'
	VAR224	'Q75-PERIOD OF NO USE-MAY 1983'

	VAR225	'Q75-PERIOD OF NO USE-NOT KNOWN'
	VAR226	'Q75-PERIOD OF NO USE-REFUSED ANSWER'
	VAR227	'Q75-PERIOD OF NO USE-NA'
	VAR228	'Q76-TYPE OF FUEL USED FOR HEATING
WATER'		
	VAR229	'Q77-SOURCE OF WATER HEATER'
	VAR230	'Q78-HOT WATER EQUIP INSTALLED AT
MOVE IN'		
	VAR231	'Q79-HOT WATER EQUIP INSTALLED-
ACTUAL'		
	VAR232	'Q80-HOT WATER EQUIP INSTALLED-ESTM'
	VAR233	'Q81-WATER HTR IN HEATED OR UNHEATED
AREA'		
	VAR234	'Q82-WATER HTR IS WRAPPED IN
INSULATION'		
	VAR235	'Q83-WRAP PROVIDED BY FUEL SUPPLIER'
	VAR236	'Q84-OTHER TYPES OF FUEL FOR WATER
HTR'		
	VAR237	'Q85-ADDITIONAL FUEL FOR WATER HTR'
	VAR238	'Q86-NUMBER OF SHOWER FACILITIES IN
HOME'		
	VAR239	'Q87-# OF SHOWERS WITH FLOW
RESTRICTORS'		
	VAR240	'Q87-# OF SHOWERS WITH LOW-FLOW
INSERTS'		
	VAR241	'Q88-# OF FAUCETS WITH LOW-FLOW
DEVICE'		
	VAR242	'Q89-PERMANENTLY REMOVED LOW-FLOW
DEVICE'		
	VAR243	'Q90-REASON FOR REMOVING LOW-FLOW
DEVICE'		
	CD6	'CARD NUMBER 6'
	ID6	'SERIAL CASE ID 6'
	VAR244	'Q91-IS HOME AIR CONDITIONED'
	VAR245	'Q92-TYPE OF AIR CONDITIONING SYSTEM'
	VAR246	'Q93-NUMBER OF AC WALL UNITS IN HOME'
	VAR247	'Q94-TYPE OF CENTRAL AC SYSTEM'
	VAR248	'Q95-USE OF AC #1'
	VAR249	'Q95-USE OF AC#2'
	VAR250	'Q96-AC#1 INSTALLED AT MOVE IN'
	VAR251	'Q96-AC#2 INSTALLED AT MOVE IN'
	VAR252	'Q97-AGE OF AIR CONDITIONER #1'
	VAR253	'Q97-AGE OF AIR CONDITIONER #2'
	VAR254	'Q98-NUMBER OF TELEVISIONS'
	VAR255	'Q99-TELEVISION #1-COLOR OR BW'
	VAR256	'Q99-TELEVISION #2-COLOR OR BW'
	VAR257	'Q100-TELEVISION #1-AGE'
	VAR258	'Q100-TELEVISION #2-AGE'
	VAR259	'Q101-TYPE OF FUEL-STOVE TOP BURNERS'
	VAR260	'Q102-STOVE TOP BURNERS INSTALLED'
	VAR261	'Q103-AGE OF STOVE TOP BURNERS'
	VAR262	'Q104-# OF OVENS USED FOR COOKING'

	VAR263	'Q105-TYPE OF FUEL FOR OVEN #1'
	VAR264	'Q105-TYPE OF FUEL FOR OVEN #2'
	VAR265	'Q106-IS OVEN #1 A MICROWAVE'
	VAR266	'Q106-IS OVEN #2 A MICROWAVE'
	VAR267	'Q107-IS OVEN #1 SEPARATE FROM TOP
BURNER '		
	VAR268	'Q107-IS OVEN #2 SEPARATE FROM TOP
BURNER '		
	VAR269	'Q108-OVEN NUMBER 1 INSTALLED'
	VAR270	'Q108-OVEN NUMBER 2 INSTALLED'
	VAR271	'Q109-AGE OF OVEN NUMBER 1'
	VAR272	'Q109-AGE OF OVEN NUMBER 2'
	VAR273	'Q110-IS OVEN #1 SELF-CLEANING'
	VAR274	'Q110-IS OVEN #2 SELF-CLEANING'
	VAR275	'Q111-NUMBER OF REFRIGERATORS IN
HOME '		
	VAR276	'Q112-SIZE OF REFRIGERATOR NUMBER 1'
	VAR277	'Q112-SIZE OF REFRIGERATOR NUMBER 2'
	VAR278	'Q113-LOCATION OF FREEZER-
REFRIGERATOR #1 '		
	VAR279	'Q113-LOCATION OF FREEZER-
REFRIGERATOR #2 '		
	VAR280	'Q114-IS REFRIGERATOR #1 FROST FREE'
	VAR281	'Q114-IS REFRIGERATOR #2 FROST FREE'
	VAR282	'Q115-HOME HAS SEPARATE FOOD FREEZER'
	VAR283	'Q116-SIZE OF SEPARATE FREEZER #1'
	VAR284	'Q116-SIZE OF SEPARATE FREEZER #2'
	VAR285	'Q117-TYPE OF SEPARATE FREEZER #1'
	VAR286	'Q117-TYPE OF SEPARATE FREEZER #2'
	VAR287	'Q118-TYPE OF FREEZER #1-FROST FREE'
	VAR288	'Q118-TYPE OF FREEZER #2-FROST FREE'
	CD7	'CARD NUMBER 7'
	ID7	'SERIAL CASE ID7'
	VAR289	'Q119-CLOTHES WASHING MACHINE-HOME
METER '		
	VAR290	'Q119-ELECTRIC CLOTHES DRYER-HOME
METER '		
	VAR291	'Q119-GAS CLOTHES DRYER-HOME METER'
	VAR292	'Q119-ELECTRIC DISHWASHER-HOME METER'
	VAR293	'Q120-DISHWASHER-SEPARATE HEATING
ELEMENT '		
	VAR294	'Q120-DISHWASHER-ENERGY SAVER SWITCH'
	VAR295	'Q121-USE ENERGY SAVER SWITCH
REGULARLY '		
	VAR296	'Q122-HEATING ELEMENT DISCONNECTED'
	VAR297	'Q123-HOME HAS SWIMMING POOL'
	VAR298	'Q123-HOME HAS HOT TUB'
	VAR299	'Q123-HOME HAS JACUZZI'
	VAR300	'Q124-FUEL USED TO HEAT POOL'
	VAR301	'Q124-FUEL USED TO HEAT HOT TUB'
	VAR302	'Q124-FUEL USED TO HEAT JACUZZI'
	VAR303	'Q125-NUMBER OF WATERBED HEATERS'

METER'	VAR304	'Q126-MEDICAL EQUIP ON HOUSEHOLD
	VAR305	'Q126-PHOTO EQUIP ON HOUSEHOLD METER'
METER'	VAR306	'Q126-WOODWORKING EQUIP ON HOUSE
	VAR307	'Q126-ELEC CERAMIC KILN ON HOUSE
METER'	VAR308	'Q126-GAS CERAMIC KILN ON HOUSE
METER'	VAR309	'Q126-HOME COMPUTER ON HOUSEHOLD
METER'	VAR310	'Q126-OFFICE EQUIP ON HOUSEHOLD
METER'	VAR311	'Q126-WELDING EQUIP ON HOUSEHOLD
METER'	VAR312	'Q126-GREENHOUSE LIGHTS ON HOUSE
METER'	VAR313	'Q126-IRRIGATION PUMPS ON HOUSEHOLD
METER'	VAR314	'Q126-ELEC WELL WATER PUMP ON HOUSE
METER'	VAR315	'Q127-OTHER EQUIP ON ELEC OR GAS
	VAR316	'Q128-OTHER EQUIP ON ELECTRIC BILL'
	VAR317	'Q128-OTHER EQUIP ON GAS BILL'
	CD8	'CARD NUMBER 8'
	ID8	'SERIAL CASE ID 8'
	VAR318	'Q129-TOTAL NUMBER OF RESIDENTS'
	VAR319	'Q130-PERSON #1-RESPONDENT CODE'
	VAR320	'Q130-PERSON #1-GENDER'
	VAR321	'Q130-PERSON #1-ACTUAL AGE'
	VAR322	'Q130-PERSON #1-AGE CATEGORY'
	VAR323	'Q130-PERSON #2-RESPONDENT CODE'
	VAR324	'Q130-PERSON #2-GENDER'
	VAR325	'Q130-PERSON #2-ACTUAL AGE'
	VAR326	'Q130-PERSON #2-AGE CATEGORY'
	VAR327	'Q130-PERSON #3-RESPONDENT CODE'
	VAR328	'Q130-PERSON #3-GENDER'
	VAR329	'Q130-PERSON #3-ACTUAL AGE'
	VAR330	'Q130-PERSON #3-AGE CATEGORY'
	VAR331	'Q130-PERSON #4-RESPONDENT CODE'
	VAR332	'Q130-PERSON #4-GENDER'
	VAR333	'Q130-PERSON #4-ACTUAL AGE'
	VAR334	'Q130-PERSON #4-AGE CATEGORY'
	VAR335	'Q130-PERSON #5-RESPONDENT CODE'
	VAR336	'Q130-PERSON #5-GENDER'
	VAR337	'Q130-PERSON #5-ACTUAL AGE'
	VAR338	'Q130-PERSON #5-AGE CATEGORY'
	VAR339	'Q130-PERSON #6-RESPONDENT CODE'
	VAR340	'Q130-PERSON #6-GENDER'
	VAR341	'Q130-PERSON #6-ACTUAL AGE'
	VAR342	'Q130-PERSON #6-AGE CATEGORY'

	VAR343	'Q130-PERSON #7-RESPONDENT CODE'
	VAR344	'Q130-PERSON #7-GENDER'
	VAR345	'Q130-PERSON #7-ACTUAL AGE'
	VAR346	'Q130-PERSON #7-AGE CATEGORY'
	VAR347	'Q130-PERSON #8-RESPONDENT CODE'
	VAR348	'Q130-PERSON #8-GENDER'
	VAR349	'Q130-PERSON #8-ACTUAL AGE'
	VAR350	'Q130-PERSON #8-AGE CATEGORY'
	VAR351	'Q130-PERSON #9-RESPONDENT CODE'
	VAR352	'Q130-PERSON #9-GENDER'
	VAR353	'Q130-PERSON #9-ACTUAL AGE'
	VAR354	'Q130-PERSON #9-AGE CATEGORY'
	VAR355	'Q130-PERSON #10-RESPONDENT CODE'
	VAR356	'Q130-PERSON #10-GENDER'
	VAR357	'Q130-PERSON #10-ACTUAL AGE'
	VAR358	'Q130-PERSON #10-AGE CATEGORY'
	VAR359	'Q131-HH#2 PARTICIPATE IN INTERVIEW'
	VAR360	'Q132-HH#1-ETHNIC ORIGIN'
	VAR361	'Q132-HH#2-ETHNIC ORIGIN'
	VAR362	'Q133-HH#1-LEVEL OF EDUCATION'
	VAR363	'Q133-HH#2-LEVEL OF EDUCATION'
	VAR364	'Q134-COMBINED 1982 INCOME'
	CD9	'CARD NUMBER 9'
	ID9	'SERIAL CASE ID9'
	VAR365	'Q135-USE OF ELECTRICITY FOR HOT
WATER'		
	VAR366	'Q136-PAYMENT FOR ELECTRICITY-HOT
WATER'		
	VAR367	'Q135-USE OF ELECTRICITY FOR HOME
HEATING'		
	VAR368	'Q136-PAYMENT FOR ELECTRICITY-HM
HEATING'		
	VAR369	'Q135-USE OF ELECTRICITY FOR AIR
COND'		
	VAR370	'Q136-PAYMENT FOR ELECTRICITY-AIR
COND'		
	VAR371	'Q135-USE OF ELECTRICITY FOR COOKING'
	VAR372	'Q136-PAYMENT FOR ELECTRICITY-
COOKING'		
	VAR373	'Q135-USE OF ELECTRICITY FOR
LIGHTING'		
	VAR374	'Q136-PAYMENT FOR ELECTRICITY-
LIGHTING'		
	VAR375	'Q135-USE OF NATURAL GAS FOR HOT
WATER'		
	VAR376	'Q136-PAYMENT FOR NATURAL GAS-HOT
WATER'		
	VAR377	'Q135-USE OF NATURAL GAS FOR HOME
HEATING'		
	VAR378	'Q136-PAYMENT FOR NATURAL GAS-HOME
HEAT'		

COND'	VAR379	'Q135-USE OF NATURAL GAS FOR AIR
COND'	VAR380	'Q136-PAYMENT FOR NATURAL GAS-AIR
COOKING'	VAR381	'Q135-USE OF NATURAL GAS FOR COOKING'
PURPOSES'	VAR382	'Q136-PAYMENT FOR NATURAL GAS-
PURPOS'	VAR383	'Q135-USE OF NATURAL GAS,MISC
	VAR384	'Q136-PAYMENT FOR NATURAL GAS-MISC
	VAR385	'Q135-USE OF TANK GAS FOR HOT WATER'
HEATING'	VAR386	'Q136-PAYMENT OF TANK GAS-HOT WATER'
HEATING'	VAR387	'Q135-USE OF TANK GAS FOR HOME
CONDITION'	VAR388	'Q136-PAYMENT FOR TANK GAS-HOME
CONDITION'	VAR389	'Q135-USE OF TANK GAS FOR AIR
	VAR390	'Q136-PAYMENT FOR TANK GAS-AIR
	VAR391	'Q135-USE OF TANK GAS FOR COOKING'
PURPOSES'	VAR392	'Q136-PAYMENT FOR TANK GAS-COOKING'
PURPOSES'	VAR393	'Q135-USE OF TANK GAS FOR MISC
	VAR394	'Q136-PAYMENT FOR TANK GAS-MISC
	VAR395	'Q135-USE OF FUEL OIL FOR HOT WATER'
HEATING'	VAR396	'Q136-PAYMENT FOR FUEL OIL-HOT WATER'
HEATING'	VAR397	'Q135-USE OF FUEL OIL FOR HOME
	VAR398	'Q136-PAYMENT FOR FUEL OIL-HOME
	VAR399L	'Q137-PIPE GAS AVAILABLE IN AREA'
	VAR400	'Q138-ELECTRICITY BILLED ON BUDGET'
	VAR401	'Q139-NATURAL GAS BILLED ON BUDGET'
	VAR402	'Q0-1-TEMPERATURE OF HOT WATER'
	VAR403	'Q0-2-HOT WATER USED IN LAST HOUR'
CLOTHES'	VAR404	'Q0-3-USED HOT WTR IN LST HR-WASH
DISHES'	VAR405	'Q0-3-USED HOT WTR IN LST HR-WASH
	VAR406	'Q0-3-USED HOT WTR IN LST HR-BATHING'
	VAR407	'Q0-3-USED HOT WTR IN LST HR-OTHER'
QUARTERS'	VAR408	'Q0-5-# OF FLOORS USED AS LIVING
	VAR409	'Q0-6-# OF FLOORS USED BELOW GROUND'
GROUND'	VAR410	'Q0-7-# OF FLOORS USED PART BELOW
	VAR411	'Q0-8-FLOOR THIS UNIT IS LOCATED'
	VAR412	'Q0-9-TYPE OF MATERIAL OUTSIDE-WOOD'
	VAR413	'Q0-9-TYPE OF MATERIAL OUTSIDE-BRICK'

	VAR414	'Q0-9-TYPE OF MATERIAL OUTSIDE-STONE'
	VAR415	'Q0-9-TYPE OF MATERIAL OUTSIDE-
CONCRETE'		
	VAR416	'Q0-9-TYPE OF MATERIAL OUTSIDE-
STUCCO'		
	VAR417	'Q0-9-TYPE OF MATERIAL OUTSIDE-ALUM
SIDE'		
	VAR418	'Q0-9-TYPE OF MATERIAL OUTSIDE-STEEL
SIDE'		
	VAR419	'Q0-9-TYPE OF MATERIAL OUTSIDE-COMP
SIDE'		
	VAR420	'Q0-9-TYPE OF MATERIAL OUTSIDE-GLASS'
	VAR421	'Q0-9-TYPE OF MATERIAL OUTSIDE-OTHER'
	CD10	'CARD NUMBER 10'
	ID10	'SERIAL CASE ID10'
	VAR422	'QB-BASEMENT TYPE'
	VAR423	'QOB-BASE UNIT A-LENGTH'
	VAR424	'QOB-BASE UNIT A-WIDTH'
	VAR425	'QOB-BASE UNIT B-LENGTH'
	VAR426	'QOB-BASE UNIT B-WIDTH'
	VAR427	'QOB-BASE UNIT C-LENGTH'
	VAR428	'QOB-BASE UNIT C-WIDTH'
	VAR429	'QOB-BASE UNIT D-LENGTH'
	VAR430	'QOB-BASE UNIT D-WIDTH'
	VAR431	'QOB-BASE UNIT X'
	VAR432	'QOB-NUMBER OF UNITS - BASEMENT'
	VAR433	'QOB-FIRST FLOOR TYPE'
	VAR434	'QOB-FIRST UNIT A-LENGTH'
	VAR435	'QOB-FIRST UNIT A-WIDTH'
	VAR436	'QOB-FIRST UNIT B-LENGTH'
	VAR437	'QOB-FIRST UNIT B-WIDTH'
	VAR438	'QOB-FIRST UNIT C-LENGTH'
	VAR439	'QOB-FIRST UNIT C-WIDTH'
	VAR440	'QOB-FIRST UNIT D-LENGTH'
	VAR441	'QOB-FIRST UNIT D-WIDTH'
	VAR442	'QOB-FIRST UNIT X'
	VAR443	'QOB-# OF UNITS'
	CD11	'CARD NUMBER 11'
	ID11	'SERIAL CASE ID11'
	VAR444	'QOB-SECOND FLOOR TYPE'
	VAR445	'QOB-SECOND FLOOR ATTIC'
	VAR446	'QOB-SECOND UNIT A-LENGTH'
	VAR447	'QOB-SECOND UNIT A-WIDTH'
	VAR448	'QOB-SECOND UNIT B-LENGTH'
	VAR449	'QOB-SECOND UNIT B-WIDTH'
	VAR450	'QOB-SECOND UNIT C-LENGTH'
	VAR451	'QOB-SECOND UNIT C-WIDTH'
	VAR452	'QOB-SECOND UNIT D-LENGTH'
	VAR453	'QOB-SECOND UNIT D-WIDTH'
	VAR454	'QOB-SECOND UNIT X'
	VAR455	'QOB-NUMBER OF UNITS - SECOND FLOOR'
	VAR456	'QOB-THIRD FLOOR TYPE'

VAR457	'QOB-THIRD FLOOR ATTIC'
VAR458	'QOB-THIRD UNIT A-LENGTH'
VAR459	'QOB-THIRD UNIT A-WIDTH'
VAR460	'QOB-THIRD UNIT B-LENGTH'
VAR461	'QOB-THIRD UNIT B-WIDTH'
VAR462	'QOB-THIRD UNIT C-LENGTH'
VAR463	'QOB-THIRD UNIT C-WIDTH'
VAR464	'QOB-THIRD UNIT D-LENGTH'
VAR465	'QOB-THIRD UNIT D-WIDTH'
VAR466	'QOB-THIRD UNIT X'
VAR467	'QOB-NUMBER OF UNITS - THIRD FLOOR'
CD12	'CARD NUMBER 12'
ID12	'SERIAL CASE ID12'
VAR468	'QOT-TOTALS BASEMENT'
VAR469	'QOT-TOTALS FIRST FLOOR'
VAR470	'QOT-TOTALS SECOND FLOOR'
VAR471	'QOT-TOTALS THIRD FLOOR'
VAR472	'QOT-TOTALS MISC.'
VAR473	'QOT-TOTALS GRAND'
VAR474	'QR1-HOW WERE MEASUREMENTS OBTAINED'
VAR475	'DATE - DAY'
VAR476	'DATE - MONTH'
VAR477	'TIME - BEGAN'
VAR478	'TIME - MORNING OR AFTERNOON'
VAR479	'TIME-END INTERVIEW-BEGIN MEASURE'
VAR480	'BEGIN MEASURE-MORNING OR AFTERNOON'
VAR481	'TIME - END MEASURE'
VAR482	'END MEASURE-MORNING OR AFTERNOON'
VAR483	'INTERVIEWER I.D.'
CD13	'CARD NUMBER 13'
ID13	'SERIAL CASE ID 13'
INDWT	'INDIVIDUAL WEIGHT'
METER	'METER READING ROUTE WEIGHT'
UTILWT	'UTILITY WEIGHT'
SUB	'SUBSTRATA WEIGHT'
STRATA	'STRATA WEIGHT'
PROJECT	'PROJECTION WEIGHT'
TOTAL	'TOTAL HOUSEHOLD WEIGHT'
ZIP	'RESIDENCE ZIP CODE'
WAIVER	'STATUS OF WAIVER FOR BILLING DATA'
NOAA	'NOAA CLIMATOLOGICAL ZONE'
FLAG	'NOAA DERIVED FROM ZIP OR ORIGINAL'

APPENDIX V

A SAMPLE OUTPUT FOR THE INFORMATION THEORETIC PROGRAM, SYSENT

This appendix presents the output for SYSENT, the program developed in order to perform the information theoretic calculations. The input consisted of cell frequencies for a contingency table of the six segmentation variables used in this study. These variables were GENATT, ENRATT, PERCEPT, AWARE, BEHAVE, and KWHUSE. The information theoretic analyses in Phase I and parts (a) and (b) of Phase II were performed using this output.

The information theoretic analyses for part (c) of Phase II were performed via a similar output, using a different contingency table. The variables for that contingency table consisted of KWHUSE, the distinguished segmentation base, and CLIMGEO, TYPDWEL, RENTOWN, DEMOG, and INSUL.

FILE: SHANON OUT

A1 PORTLAND STATE UNIVERSITY VM/SP R4 03/27/87

NUMBER	ENTPOP	TRANSMISSION	SYS. ENTRGP	0=ABSNT, 1=PRES, 2=GIVEN
1	0.00000000	0.00000000	0.00000000	0 0 0 0 0 0
2	1.55713463	0.00000000	0.00000000	1 0 0 0 0 0
3	0.00000000	0.00000000	0.00000000	2 0 0 0 0 0
4	2.24130917	0.00000000	0.00000000	0 1 0 0 0 0
5	3.79064655	0.00779724	0.00779724	1 1 0 0 0 0
6	2.23351192	0.00000000	0.00000000	2 1 0 0 0 0
7	0.00000000	0.00000000	0.00000000	0 2 0 0 0 0
8	1.54933739	0.00000000	0.00000000	1 2 0 0 0 0
9	0.00000000	0.00000000	0.00000000	2 2 0 0 0 0
10	1.36871147	0.00000000	0.00000000	0 0 1 0 0 0
11	2.88244820	0.04339790	0.04339790	1 0 1 0 0 0
12	1.32531357	0.00000000	0.00000000	2 0 1 0 0 0
13	3.60473251	0.00528812	0.00528812	0 1 1 0 0 0
14	5.10079575	0.06635952	0.07623577	1 1 1 0 0 0
15	3.54366112	0.01516438	0.01516438	2 1 1 0 0 0
16	1.36342335	0.00000000	0.00000000	0 2 1 0 0 0
17	2.85948653	0.05327415	0.05327415	1 2 1 0 0 0
18	1.31014919	0.00000000	0.00000000	2 2 1 0 0 0
19	0.00000000	0.00000000	0.00000000	0 0 2 0 0 0
20	1.51373672	0.00000000	0.00000000	1 0 2 0 0 0
21	0.00000000	0.00000000	0.00000000	2 0 2 0 0 0
22	2.23602104	0.00000000	0.00000000	0 1 2 0 0 0
23	3.73208427	0.01767349	0.01767349	1 1 2 0 0 0
24	2.21834755	0.00000000	0.00000000	2 1 2 0 0 0
25	0.00000000	0.00000000	0.00000000	0 2 2 0 0 0
26	1.49606323	0.00000000	0.00000000	1 2 2 0 0 0
27	0.00000000	0.00000000	0.00000000	2 2 2 0 0 0
28	1.01307964	0.00000000	0.00000000	0 0 0 1 0 0
29	2.56782722	0.00238705	0.00238705	1 0 0 1 0 0
30	1.01069260	0.00000000	0.00000000	2 0 0 1 0 0
31	3.24918652	0.00520229	0.00520229	0 1 0 1 0 0
32	4.79102421	0.02049923	0.02561138	1 1 0 1 0 0
33	3.23388958	0.01031494	0.01031494	2 1 0 1 0 0
34	1.00787735	0.00000000	0.00000000	0 2 0 1 0 0
35	2.54971504	0.00749969	0.00749969	1 2 0 1 0 0
36	1.00037766	0.00000000	0.00000000	2 2 0 1 0 0
37	2.37683487	0.00495625	0.00495625	0 0 1 1 0 0
38	3.38514614	0.05377960	0.05681801	1 0 1 1 0 0
39	2.32801151	0.00799465	0.00799465	2 0 1 1 0 0
40	4.60252053	0.02017975	0.02491283	0 1 1 1 0 0
41	6.07762146	0.10261345	0.14702225	1 1 1 1 0 0
42	4.52048683	0.04903126	0.06458855	2 1 1 1 0 0
43	2.35161137	0.00968933	0.00968933	0 2 1 1 0 0
44	3.63631229	0.05432579	0.09318840	1 2 1 1 0 0
45	2.28697491	0.02355194	0.02355194	2 2 1 1 0 0
46	1.00812340	0.00000000	0.00000000	0 0 2 1 0 0
47	2.51643467	0.00542545	0.00542545	1 0 2 1 0 0
48	1.00269794	0.00000000	0.00000000	2 0 2 1 0 0
49	3.23420906	0.00993538	0.00993538	0 1 2 1 0 0
50	4.70890999	0.04897118	0.06490303	1 1 2 1 0 0
51	3.19517326	0.02587223	0.02587223	2 1 2 1 0 0
52	0.99218902	0.00000000	0.00000000	0 2 2 1 0 0
53	2.47233895	0.02136230	0.02136230	1 2 2 1 0 0

FILE: SHANON OUT A1 PORTLAND STATE UNIVERSITY VM/SP R4 03/27/87 .

54	0.97682571	0.00000000	0.00000000	2	2	2	1	0	0
55	0.00000000	0.00000000	0.00000000	0	0	0	2	0	0
56	1.55474758	0.00000000	0.00000000	1	0	0	2	0	0
57	0.00000000	0.00000000	0.00000000	2	0	0	2	0	0
58	2.23610687	0.00000000	0.00000000	0	1	0	2	0	0
59	3.77794456	0.01290989	0.01290989	1	1	0	2	0	0
60	2.22319698	0.00000000	0.00000000	2	1	0	2	0	0
61	0.00000000	0.00000000	0.00000000	0	2	0	2	0	0
62	1.54183769	0.00000000	0.00000000	1	2	0	2	0	0
63	0.00000000	0.00000000	0.00000000	2	2	0	2	0	0
64	1.36375523	0.00000000	0.00000000	0	0	1	2	0	0
65	2.87206650	0.04643631	0.04643631	1	0	1	2	0	0
66	1.31731892	0.00000000	0.00000000	2	0	1	2	0	0
67	3.58984089	0.01002121	0.01002121	0	1	1	2	0	0
68	5.06454182	0.09006786	0.11076832	1	1	1	2	0	0
69	3.50979424	0.03072166	0.03072166	2	1	1	2	0	0
70	1.35373402	0.00000000	0.00000000	0	2	1	2	0	0
71	2.82843494	0.06713676	0.06713676	1	2	1	2	0	0
72	1.28559725	0.00000000	0.00000000	2	2	1	2	0	0
73	0.00000000	0.00000000	0.00000000	0	0	2	2	0	0
74	1.50831127	0.00000000	0.00000000	1	0	2	2	0	0
75	0.00000000	0.00000000	0.00000000	2	0	2	2	0	0
76	2.22508566	0.00000000	0.00000000	0	1	2	2	0	0
77	3.70078659	0.03361034	0.03361034	1	1	2	2	0	0
78	2.19247532	0.00000000	0.00000000	2	1	2	2	0	0
79	0.00000000	0.00000000	0.00000000	0	2	2	2	0	0
80	1.47470093	0.00000000	0.00000000	1	2	2	2	0	0
81	0.00000000	0.00000000	0.00000000	2	2	2	2	0	0
82	0.61858737	0.00000000	0.00000000	0	0	0	0	1	0
83	2.17317677	0.00254524	0.00254536	1	0	0	0	1	0
84	0.61604214	0.00000000	0.00000000	2	0	0	0	1	0
85	2.85634708	0.00354946	0.00354958	0	1	0	0	1	0
86	4.40094852	0.01608264	0.01827335	1	1	0	0	1	0
87	2.84381390	0.00574017	0.00574017	2	1	0	0	1	0
88	0.61503792	0.00000000	0.00000000	0	2	0	0	1	0
89	2.15963936	0.00473595	0.00473595	1	2	0	0	1	0
90	0.61030197	0.00000000	0.00000000	2	2	0	0	1	0
91	1.98446846	0.00283039	0.00283051	0	0	1	0	1	0
92	3.49426937	0.05016410	0.05155468	1	0	1	0	1	0
93	1.93713474	0.00422096	0.00422096	2	0	1	0	1	0
94	4.21551609	0.01309192	0.01451588	0	1	1	0	1	0
95	5.70090103	0.08484161	0.10882664	1	1	1	0	1	0
96	4.14376640	0.03110123	0.03707695	2	1	1	0	1	0
97	1.97420692	0.00425434	0.00425434	0	2	1	0	1	0
98	3.45959187	0.06820679	0.07414913	1	2	1	0	1	0
99	1.91025448	0.01019669	0.01019669	2	2	1	0	1	0
100	0.61575699	0.00000000	0.00000000	0	0	2	0	1	0
101	2.12555790	0.00393581	0.00393581	1	0	2	0	1	0
102	0.61182117	0.00000000	0.00000000	2	0	2	0	1	0
103	2.84680462	0.00497341	0.00497341	0	1	2	0	1	0
104	4.33218956	0.03332520	0.04006767	1	1	2	0	1	0
105	2.81945284	0.01171589	0.01171589	2	1	2	0	1	0
106	0.61078358	0.00000000	0.00000000	0	2	2	0	1	0
107	2.09616852	0.01067829	0.01067829	1	2	2	0	1	0
108	0.60010529	0.00000000	0.00000000	2	2	2	0	1	0

FILE: SHANON OUT A1 PORTLAND STATE UNIVERSITY VM/SP R4 03/27/87 .

109	1.23816204	0.39350493	0.39350498	0	0	0	1	1	0
110	2.78973866	0.39906299	0.39968572	1	0	0	1	1	0
111	1.23260403	0.39413071	0.39413071	2	0	0	1	1	0
112	3.47102642	0.40194976	0.40164280	0	1	0	1	1	0
113	5.00347042	0.42664039	0.44232655	1	1	0	1	1	0
114	3.44633579	0.41391087	0.41763592	2	1	0	1	1	0
115	1.22971725	0.39319801	0.39319801	0	2	0	1	1	0
116	2.76216125	0.41009140	0.41474915	1	2	0	1	1	0
117	1.21282387	0.39785576	0.39785576	2	2	0	1	1	0
118	2.60043621	0.39994228	0.39859295	0	0	1	1	1	0
119	4.10234737	0.45516574	0.46254826	1	0	1	1	1	0
120	2.54521275	0.40683556	0.40732479	2	0	1	1	1	0
121	4.81755161	0.42413604	0.43724442	0	1	1	1	1	0
122	6.27415848	0.52466381	0.60525799	1	1	1	1	1	0
123	4.71702385	0.46853638	0.50473022	2	1	1	1	1	0
124	2.57624245	0.41009617	0.41305065	0	2	1	1	1	0
125	4.03284931	0.50282669	0.53575993	1	2	1	1	1	0
126	2.48351192	0.43731689	0.44302940	2	2	1	1	1	0
127	1.23172474	0.39215565	0.39215565	0	0	2	1	1	0
128	2.73363590	0.40398121	0.40644550	1	0	2	1	1	0
129	1.21989918	0.39461994	0.39461994	2	0	2	1	1	0
130	3.44884014	0.41106129	0.41505814	0	1	2	1	1	0
131	4.90544701	0.46819115	0.50723457	1	1	2	1	1	0
132	3.39171028	0.44115639	0.45010471	2	1	2	1	1	0
133	1.21281910	0.39615250	0.39615250	0	2	2	1	1	0
134	2.66942596	0.43560836	0.44302464	1	2	2	1	1	0
135	1.17336273	0.40356827	0.40356827	2	2	2	1	1	0
136	0.22508240	0.00000000	0.00000000	0	0	0	2	1	0
137	1.77665901	0.00317097	0.00317097	1	0	0	2	1	0
138	0.22191143	0.00000000	0.00000000	2	0	0	2	1	0
139	2.45794678	0.00324249	0.00324249	0	1	0	2	1	0
140	3.99039078	0.02554607	0.03176880	1	1	0	2	1	0
141	2.43564320	0.00946522	0.00946522	2	1	0	2	1	0
142	0.22183990	0.00000000	0.00000000	0	2	0	2	1	0
143	1.75428391	0.00939369	0.00939369	1	2	0	2	1	0
144	0.21244621	0.00000000	0.00000000	2	2	0	2	1	0
145	1.58735657	0.00148106	0.00148106	0	0	1	2	1	0
146	3.08926773	0.05431747	0.05754662	1	0	1	2	1	0
147	1.53452015	0.00471020	0.00471020	2	0	1	2	1	0
148	3.80447197	0.02047253	0.02620029	0	1	1	2	1	0
149	5.26107983	0.11361324	0.16543579	1	1	1	2	1	0
150	3.70633125	0.05609608	0.06729507	2	1	1	2	1	0
151	1.56836510	0.00720882	0.00720882	0	2	1	2	1	0
152	3.02497196	0.09243965	0.10114002	1	2	1	2	1	0
153	1.48313427	0.01590919	0.01590919	2	2	1	2	1	0
154	0.22360134	0.00000000	0.00000000	0	0	2	2	1	0
155	1.72551250	0.00640011	0.00640011	1	0	2	2	1	0
156	0.21720123	0.00000000	0.00000000	2	0	2	2	1	0
157	2.44071674	0.00897020	0.00897020	0	1	2	2	1	0
158	3.89732361	0.06067467	0.07236862	1	1	2	2	1	0
159	2.38901234	0.02066422	0.02066422	2	1	2	2	1	0
160	0.21463103	0.00000000	0.00000000	0	2	2	2	1	0
161	1.67123795	0.01809406	0.01809406	1	2	2	2	1	0
162	0.19653702	0.00000000	0.00000000	2	2	2	2	1	0
163	0.00000000	0.00000000	0.00000000	0	0	0	0	2	0

FILE: SHANON OUT A1 PORTLAND STATE UNIVERSITY VM/SP R4 03/27/87

164	1.55458927	0.00000000	0.00000000	1	0	0	0	2	0
165	0.00000000	0.00000000	0.00000000	2	0	0	0	2	0
166	2.23775959	0.00000000	0.00000000	0	1	0	0	2	0
167	3.79236103	0.00998783	0.00998783	1	1	0	0	2	0
168	2.22777176	0.00000000	0.00000000	2	1	0	0	2	0
169	0.00000000	0.00000000	0.00000000	0	2	0	0	2	0
170	1.54460144	0.00000000	0.00000000	1	2	0	0	2	0
171	0.00000000	0.00000000	0.00000000	2	2	0	0	2	0
172	1.36589097	0.00000000	0.00000000	0	0	1	0	2	0
173	2.87568188	0.04478836	0.04478836	1	0	1	0	2	0
174	1.32109261	0.00000000	0.00000000	2	0	1	0	2	0
175	3.59692860	0.00671196	0.00671196	0	1	1	0	2	0
176	5.09231354	0.07591629	0.07591629	1	1	1	0	2	0
177	3.52772427	0.02114010	0.02114010	2	1	1	0	2	0
178	1.35916901	0.00000000	0.00000000	0	2	1	0	2	0
179	2.84455395	0.05921650	0.05921650	1	2	1	0	2	0
180	1.29995251	0.00000000	0.00000000	2	2	1	0	2	0
181	0.00000000	0.00000000	0.00000000	0	0	2	0	2	0
182	1.50980091	0.00000000	0.00000000	1	0	2	0	2	0
183	0.00000000	0.00000000	0.00000000	2	0	2	0	2	0
184	2.23104763	0.00000000	0.00000000	0	1	2	0	2	0
185	3.71643257	0.02441597	0.02441597	1	1	2	0	2	0
186	2.20663166	0.00000000	0.00000000	2	1	2	0	2	0
187	0.00000000	0.00000000	0.00000000	0	2	2	0	2	0
188	1.49538494	0.00000000	0.00000000	1	2	2	0	2	0
189	0.00000000	0.00000000	0.00000000	2	2	2	0	2	0
190	0.61957467	0.00000000	0.00000000	0	0	0	1	2	0
191	2.17115116	0.00301278	0.00301278	1	0	0	1	2	0
192	0.61656189	0.00000000	0.00000000	2	0	0	1	2	0
193	2.85243893	0.00489533	0.00489533	0	1	0	1	2	0
194	4.32488293	0.02704060	0.02704060	1	1	0	1	2	0
195	2.83029366	0.01403999	0.01403999	2	1	0	1	2	0
196	0.61467934	0.00000000	0.00000000	0	2	0	1	2	0
197	2.14712334	0.01215744	0.01215744	1	2	0	1	2	0
198	0.60252190	0.00000000	0.00000000	2	2	0	1	2	0
199	1.98184872	0.00360692	0.00360692	0	0	1	1	2	0
200	3.49375988	0.05628502	0.05628502	1	0	1	1	2	0
201	1.92917061	0.00848389	0.00848389	2	0	1	1	2	0
202	4.19896412	0.02425110	0.02425110	0	1	1	1	2	0
203	5.65557098	0.12223351	0.12223351	1	1	1	1	2	0
204	4.10098171	0.06444454	0.06444454	2	1	1	1	2	0
205	1.96120453	0.01264381	0.01264381	0	2	1	1	2	0
206	3.41781139	0.10063839	0.10063839	1	2	1	1	2	0
207	1.87320995	0.02926445	0.02926445	2	2	1	1	2	0
208	0.61596775	0.00000000	0.00000000	0	0	2	1	2	0
209	2.11787891	0.00788975	0.00788975	1	0	2	1	2	0
210	0.60907900	0.00000000	0.00000000	2	0	2	1	2	0
211	2.83308315	0.01393223	0.01393223	0	1	2	1	2	0
212	4.28969002	0.06712627	0.06712627	1	1	2	1	2	0
213	2.77988911	0.03482056	0.03482056	2	1	2	1	2	0
214	0.60203552	0.00000000	0.00000000	0	2	2	1	2	0
215	2.05864239	0.02877808	0.02877808	1	2	2	1	2	0
216	0.57325745	0.00000000	0.00000000	2	2	2	1	2	0
217	0.00000000	0.00000000	0.00000000	0	0	0	2	2	0
218	1.55157661	0.00000000	0.00000000	1	0	0	2	2	0

FILE: SHANON QUT A1 PORTLAND STATE UNIVERSITY VM/SP R4 03/27/87 .

219	0.00000000	0.00000000	0.00000000	2	0	0	2	2	0
220	2.23286438	0.00000000	0.00000000	0	1	0	2	2	0
221	3.76530838	0.01913261	0.01913261	1	1	0	2	2	0
222	2.21373177	0.00000000	0.00000000	2	1	0	2	2	0
223	0.00000000	0.00000000	0.00000000	0	2	0	2	2	0
224	1.53244400	0.00000000	0.00000000	1	2	0	2	2	0
225	0.00000000	0.00000000	0.00000000	2	2	0	2	2	0
226	1.36227417	0.00000000	0.00000000	0	0	1	2	2	0
227	2.86418533	0.04966545	0.04966545	1	0	1	2	2	0
228	1.31260872	0.00000000	0.00000000	2	0	1	2	2	0
229	3.57938957	0.01574898	0.01574898	0	1	1	2	2	0
230	5.03599644	0.11071873	0.13689041	1	1	1	2	2	0
231	3.48441982	0.04192066	0.04192066	2	1	1	2	2	0
232	1.34652519	0.00000000	0.00000000	0	2	1	2	2	0
233	2.80313206	0.07583714	0.07583714	1	2	1	2	2	0
234	1.27068806	0.00000000	0.00000000	2	2	1	2	2	0
235	0.00000000	0.00000000	0.00000000	0	0	2	2	2	0
236	1.50191116	0.00000000	0.00000000	1	0	2	2	2	0
237	0.00000000	0.00000000	0.00000000	2	0	2	2	2	0
238	2.21711540	0.00000000	0.00000000	0	1	2	2	2	0
239	3.67372227	0.04530430	0.04530430	1	1	2	2	2	0
240	2.17181110	0.00000000	0.00000000	2	1	2	2	2	0
241	0.00000000	0.00000000	0.00000000	0	2	2	2	2	0
242	1.45660686	0.00000000	0.00000000	1	2	2	2	2	0
243	0.00000000	0.00000000	0.00000000	2	2	2	2	2	0
244	1.42557621	0.00000000	0.00000000	0	0	0	0	0	1
245	2.98135567	0.00135517	0.00135517	1	0	0	0	0	1
246	1.42422104	0.00000000	0.00000000	2	0	0	0	0	1
247	3.65924931	0.00763607	0.00763607	0	1	0	0	0	1
248	5.19410133	0.02991867	0.04304886	1	1	0	0	0	1
249	3.63696671	0.02076626	0.02076626	2	1	0	0	0	1
250	1.41794014	0.00000000	0.00000000	0	2	0	0	0	1
251	2.95279217	0.01448536	0.01448536	1	2	0	0	0	1
252	1.40345478	0.00000000	0.00000000	2	2	0	0	0	1
253	2.78937395	0.00591373	0.00591373	0	0	1	0	0	1
254	4.29417515	0.05724716	0.06382751	1	0	1	0	0	1
255	2.73704052	0.01249409	0.01249409	2	0	1	0	0	1
256	5.00633049	0.02926636	0.03969479	0	1	1	0	0	1
257	6.46603584	0.12669563	0.19729519	1	1	1	0	0	1
258	4.90890121	0.07414532	0.09986591	2	1	1	0	0	1
259	2.76502132	0.01634216	0.01634216	0	2	1	0	0	1
260	4.22472668	0.10597420	0.12784672	1	2	1	0	0	1
261	2.67538929	0.03821468	0.03821468	2	2	1	0	0	1
262	1.41966248	0.00000000	0.00000000	0	0	2	0	0	1
263	2.92546368	0.00793552	0.00793552	1	0	2	0	0	1
264	1.41172695	0.00000000	0.00000000	2	0	2	0	0	1
265	3.63761902	0.01806450	0.01806450	0	1	2	0	0	1
266	5.09732437	0.07209587	0.10051823	1	1	2	0	0	1
267	3.58358765	0.04648685	0.04648685	2	1	2	0	0	1
268	1.40159798	0.00000000	0.00000000	0	2	2	0	0	1
269	2.86130333	0.03635788	0.03635788	1	2	2	0	0	1
270	1.36524010	0.00000000	0.00000000	2	2	2	0	0	1
271	2.43802547	0.00063038	0.00063038	0	0	0	1	0	1
272	3.98556995	0.01022053	0.01606846	1	0	0	1	0	1
273	2.42843533	0.00647831	0.00647831	2	0	0	1	0	1

FILE: SHANON OUT

A1 PORTLAND STATE UNIVERSITY VM/SP R4 03/27/87

274	4.65892874	0.02113628	0.02880383	0	1	0	1	0	1
275	6.16521263	0.07188702	0.13388634	1	1	0	1	0	1
276	4.60807600	0.06034755	0.08313560	2	1	0	1	0	1
277	2.41751957	0.00629792	0.00829792	0	2	0	1	0	1
278	3.92390347	0.05125141	0.07221985	1	2	0	1	0	1
279	2.37456606	0.02926636	0.02926636	2	2	0	1	0	1
280	3.79123592	0.01613140	0.02076244	0	0	1	1	0	1
281	5.27728558	0.08721638	0.12427044	1	0	1	1	0	1
282	3.72015095	0.04007626	0.05318546	2	0	1	1	0	1
283	5.97898197	0.06969452	0.12236977	0	1	1	1	0	1
284	7.28121510	0.22459602	0.44027710	1	1	1	1	0	1
285	5.82408047	0.16965866	0.28537560	2	1	1	1	0	1
286	3.73767281	0.05156803	0.06880665	0	2	1	1	0	1
287	5.13990593	0.19867229	0.30289745	1	2	1	1	0	1
288	3.59056854	0.12341309	0.15579319	2	2	1	1	0	1
289	2.42252445	0.00526142	0.00526142	0	0	2	1	0	1
290	3.90857410	0.03294849	0.04727459	1	0	2	1	0	1
291	2.39483738	0.01958752	0.01958752	2	0	2	1	0	1
292	4.61027050	0.05353642	0.07381153	0	1	2	1	0	1
293	6.01250362	0.16504002	0.28756809	1	1	2	1	0	1
294	4.49876690	0.13400555	0.17606449	2	1	2	1	0	1
295	2.37424946	0.02553654	0.02553654	0	2	2	1	0	1
296	3.77648258	0.11936665	0.15547657	1	2	2	1	0	1
297	2.28041935	0.06164646	0.06164646	2	2	2	1	0	1
298	1.42494583	0.00000000	0.00000000	0	0	0	2	0	1
299	2.97249031	0.00720310	0.00720310	1	0	0	2	0	1
300	1.41774273	0.00000000	0.00000000	2	0	0	2	0	1
301	3.64574909	0.01530361	0.01530361	0	1	0	2	0	1
302	5.15213299	0.06366730	0.09191799	1	1	0	2	0	1
303	3.59738541	0.04355431	0.04355431	2	1	0	2	0	1
304	1.40964222	0.00000000	0.00000000	0	2	0	2	0	1
305	2.91602612	0.03545380	0.03545380	1	2	0	2	0	1
306	1.37418842	0.00000000	0.00000000	2	2	0	2	0	1
307	2.77815628	0.01054478	0.01054478	0	0	1	2	0	1
308	4.26420593	0.07924271	0.09430122	1	0	1	2	0	1
309	2.70945835	0.02560329	0.02560329	2	0	1	2	0	1
310	4.96590233	0.05890560	0.08194160	0	1	1	2	0	1
311	6.36813545	0.21142006	0.34237671	1	1	1	2	0	1
312	4.81338787	0.14487076	0.18986225	2	1	1	2	0	1
313	2.72979546	0.03358078	0.03358078	0	2	1	2	0	1
314	4.13202858	0.17318535	0.21019936	1	2	1	2	0	1
315	2.59019089	0.07059479	0.07059479	2	2	1	2	0	1
316	1.41440105	0.00000000	0.00000000	0	0	2	2	0	1
317	2.90045071	0.02226162	0.02226162	1	0	2	2	0	1
318	1.39213943	0.00000000	0.00000000	2	0	2	2	0	1
319	3.60214710	0.03833961	0.03833961	0	1	2	2	0	1
320	5.00438023	0.14441776	0.19462395	1	1	2	2	0	1
321	3.49606895	0.08854580	0.08854580	2	1	2	2	0	1
322	1.37606144	0.00000000	0.00000000	0	2	2	2	0	1
323	2.77829456	0.07246780	0.07246780	1	2	2	2	0	1
324	1.30359364	0.00000000	0.00000000	2	2	2	2	0	1
325	2.04202090	0.00214291	0.00214291	0	0	0	0	1	1
326	3.59194374	0.00935459	0.01266575	1	0	0	0	1	1
327	2.03480911	0.00545406	0.00545406	2	0	0	0	1	1
328	4.26909924	0.01637363	0.01941872	0	1	0	0	1	1

FILE= SHANON OUT A1 PORTLAND STATE UNIVERSITY VM/SP R4 03/27/87

329	5.79047775	0.05212975	0.08465958	1	1	0	0	1	1
330	4.23334312	0.04043198	0.04890347	2	1	0	0	1	1
331	2.02779007	0.00518799	0.00518799	0	2	0	0	1	1
332	3.54916859	0.03314686	0.04188442	1	2	0	0	1	1
333	1.99983120	0.01392555	0.01392555	2	2	0	0	1	1
334	3.39881134	0.01406384	0.01724052	0	0	1	0	1	1
335	4.89359283	0.07641697	0.09842110	1	0	1	0	1	1
336	3.33645821	0.02911854	0.03606796	2	0	1	0	1	1
337	5.60275555	0.05142879	0.08149052	0	1	1	0	1	1
338	7.03193092	0.17938805	0.32603931	1	1	1	0	1	1
339	5.47479630	0.12429237	0.19808006	2	1	1	0	1	1
340	3.35144638	0.03495502	0.04412556	0	2	1	0	1	1
341	4.79062176	0.15511703	0.22306824	1	2	1	0	1	1
342	3.24128437	0.08262157	0.10290623	2	2	1	0	1	1
343	2.03009987	0.00531960	0.00531960	0	0	2	0	1	1
344	3.52488136	0.02427483	0.03135872	1	0	2	0	1	1
345	2.01114464	0.01240349	0.01240349	2	0	2	0	1	1
346	4.23404408	0.03739643	0.04643536	0	1	2	0	1	1
347	5.66321945	0.12195773	0.19878101	1	1	2	0	1	1
348	4.14948273	0.09241295	0.11421967	2	1	2	0	1	1
349	1.99802303	0.01435852	0.01435852	0	2	2	0	1	1
350	3.42719841	0.08124638	0.10109806	1	2	2	0	1	1
351	1.93113518	0.03421021	0.03421021	2	2	2	0	1	1
352	2.65618420	0.40105915	0.40583992	0	0	0	1	1	1
353	4.19594765	0.41843033	0.43559361	1	0	0	1	1	1
354	2.63381302	0.41214275	0.41822243	2	0	0	1	1	1
355	4.86338436	0.43016803	0.44998550	0	1	0	1	1	1
356	6.36062527	0.49506176	0.58099174	1	1	0	1	1	1
357	4.80349064	0.48097706	0.51609802	2	1	0	1	1	1
358	2.62707520	0.41378021	0.42087650	0	2	0	1	1	1
359	4.11931610	0.47087669	0.50436020	1	2	0	1	1	1
360	2.56997971	0.44415569	0.44726372	2	2	0	1	1	1
361	4.00352383	0.42243099	0.43609619	0	0	1	1	1	1
362	5.47677598	0.50631034	0.56559132	1	0	1	1	1	1
363	3.91964436	0.45662498	0.48170185	2	0	1	1	1	1
364	6.17160034	0.49566352	0.58479595	0	1	1	1	1	1
365	7.54303837	0.68136013	0.98111725	1	1	1	1	1	1
366	5.98590374	0.62387753	0.79542065	2	1	1	1	1	1
367	3.93029118	0.47398758	0.51156330	0	2	1	1	1	1
368	5.30172920	0.65188694	0.80606747	1	2	1	1	1	1
369	3.75239182	0.57189178	0.62816811	2	2	1	1	1	1
370	2.63481236	0.40873051	0.41472435	0	0	2	1	1	1
371	4.10806751	0.44721207	0.47770119	1	0	2	1	1	1
372	2.59433079	0.43191528	0.43721962	2	0	2	1	1	1
373	4.80298887	0.47667503	0.51930046	0	1	2	1	1	1
374	6.17432690	0.61897373	0.79481888	1	1	2	1	1	1
375	4.66059017	0.58400345	0.65252018	2	1	2	1	1	1
376	2.56686793	0.44370174	0.45135593	0	2	2	1	1	1
377	3.93830585	0.56832695	0.62505722	1	2	2	1	1	1
378	2.44224262	0.49992847	0.50043201	2	2	2	1	1	1
379	1.64310455	0.00692368	0.00692368	0	0	0	2	1	1
380	3.19286800	0.02190781	0.02651787	1	0	0	2	1	1
381	1.62312042	0.01153374	0.01153374	2	0	0	2	1	1
382	3.85530472	0.03083033	0.03619099	0	1	0	2	1	1
383	5.34754562	0.09333706	0.13805962	1	1	0	2	1	1

FILE: SHANCN OUT A1 PORTLAND STATE UNIVERSITY VM/SP R4 03/27/87

384	3.79279804	0.07005310	0.07555294	2	1	0	2	1	1
385	1.61919735	0.01228428	0.01228428	0	2	0	2	1	1
386	3.11143875	0.06138107	0.06663036	1	2	0	2	1	1
387	1.56960106	0.01703358	0.01703353	2	2	0	2	1	1
388	2.99044418	0.02333927	0.02772903	0	0	1	2	1	1
389	4.46369934	0.10483170	0.13568783	1	0	1	2	1	1
390	2.90895176	0.04302132	0.05419540	2	0	1	2	1	1
391	5.15852070	0.09136963	0.14056110	0	1	1	2	1	1
392	6.52995872	0.27467918	0.47914505	1	1	1	2	1	1
393	4.97521114	0.20495892	0.29583549	2	1	1	2	1	1
394	2.92241383	0.06280231	0.07253075	0	2	1	2	1	1
395	4.29385185	0.23320198	0.30929756	1	2	1	2	1	1
396	2.75201416	0.12121773	0.13889790	2	2	1	2	1	1
397	1.62669896	0.01131344	0.01131344	0	0	2	2	1	1
398	3.09994411	0.04636955	0.05276394	1	0	2	2	1	1
399	1.59163284	0.01770782	0.01770782	2	0	2	2	1	1
400	3.79476547	0.06932259	0.08002186	0	1	2	2	1	1
401	5.16620350	0.20619583	0.29780293	1	1	2	2	1	1
402	3.65789223	0.14392376	0.16092968	2	1	2	2	1	1
403	1.56867981	0.02201271	0.02201271	0	2	2	2	1	1
404	2.94011784	0.12527561	0.13797665	1	2	2	2	1	1
405	1.46541691	0.03471375	0.03471375	2	2	2	2	1	1
406	1.42343330	0.00000000	0.00000000	0	0	0	0	2	1
407	2.97335625	0.00466633	0.00466633	1	0	0	0	2	1
408	1.41876598	0.00000000	0.00000000	2	0	0	0	2	1
409	3.65051174	0.01068115	0.01068115	0	1	0	0	2	1
410	5.17189026	0.04389191	0.06244850	1	1	0	0	2	1
411	3.61730099	0.02923775	0.02923775	2	1	0	0	2	1
412	1.41275215	0.00000000	0.00000000	0	2	0	0	2	1
413	2.93413067	0.02322292	0.02322292	1	2	0	0	2	1
414	1.38952923	0.00000000	0.00000000	2	2	0	0	2	1
415	2.79022385	0.00909042	0.00909042	0	0	1	0	2	1
416	4.27500534	0.05889820	0.07925129	1	0	1	0	2	1
417	2.72041607	0.01944351	0.01944351	2	0	1	0	2	1
418	4.98416805	0.04290581	0.05932803	0	1	1	0	2	1
419	6.41334343	0.16831970	0.27334690	1	1	1	0	2	1
420	4.85375416	0.10887718	0.14793301	2	1	1	0	2	1
421	2.74640846	0.02551270	0.02551270	0	2	1	0	2	1
422	4.17558384	0.14093876	0.17392540	1	2	1	0	2	1
423	2.63098240	0.05849934	0.05849934	2	2	1	0	2	1
424	1.41434288	0.00000000	0.00000000	0	0	2	0	2	1
425	2.90912437	0.01501942	0.01501942	1	0	2	0	2	1
426	1.39932346	0.00000000	0.00000000	2	0	2	0	2	1
427	3.61828709	0.02710342	0.02710342	0	1	2	0	2	1
428	5.04746246	0.10772896	0.14891911	1	1	2	0	2	1
429	3.53766155	0.06829357	0.06829357	2	1	2	0	2	1
430	1.38723946	0.00000000	0.00000000	0	2	2	0	2	1
431	2.81641483	0.05620956	0.05620956	1	2	2	0	2	1
432	1.33102989	0.00000000	0.00000000	2	2	2	0	2	1
433	2.03759670	0.00541210	0.00541115	0	0	0	1	2	1
434	3.57736015	0.02023792	0.02738380	1	0	0	1	2	1
435	2.02277088	0.01255793	0.01255798	2	0	0	1	2	1
436	4.24979687	0.03097153	0.04095364	0	1	0	1	2	1
437	5.74203777	0.09331989	0.15781689	1	1	0	1	2	1
438	4.18744850	0.07565212	0.09546852	2	1	0	1	2	1

FILE: SHANON OUT A1 PORTLAND STATE UNIVERSITY VM/SP R4 03/27/87

439	2.01203723	0.01539421	0.01539421	0	2	0	1	2	1
440	3.50427818	0.06775475	0.08473492	1	2	0	1	2	1
441	1.95967674	0.03237438	0.03237438	2	2	0	1	2	1
442	3.38493633	0.02395344	0.02979660	0	0	1	1	2	1
443	4.85819149	0.10528755	0.14648724	1	0	1	1	2	1
444	3.30360222	0.05281925	0.06515312	2	0	1	1	2	1
445	5.55301285	0.09363651	0.15882683	0	1	1	1	2	1
446	6.92445087	0.27678776	0.52435303	1	1	1	1	2	1
447	5.36986160	0.21433163	0.34120178	2	1	1	1	2	1
448	3.31525326	0.07134724	0.08914375	0	2	1	1	2	1
449	4.68669128	0.24451065	0.35285282	1	2	1	1	2	1
450	3.14208984	0.14991379	0.17968941	2	2	1	1	2	1
451	2.01905537	0.01125526	0.01125526	0	0	2	1	2	1
452	3.49231052	0.04780102	0.06143761	1	0	2	1	2	1
453	1.98250961	0.02489185	0.02489185	2	0	2	1	2	1
454	4.18713188	0.07422638	0.09616184	0	1	2	1	2	1
455	5.55856991	0.21258926	0.34088516	1	1	2	1	2	1
456	4.04876900	0.16526413	0.20252228	2	1	2	1	2	1
457	1.95603425	0.03319073	0.03319073	0	2	2	1	2	1
458	3.32752228	0.14713764	0.17609692	1	2	2	1	2	1
459	1.84213734	0.06215000	0.06215000	2	2	2	1	2	1
460	1.41802216	0.00000000	0.00000000	0	0	0	2	2	1
461	2.95778561	0.01181316	0.01181316	1	0	0	2	2	1
462	1.40620399	0.00000000	0.00000000	2	0	0	2	2	1
463	3.63022232	0.02066422	0.02066422	0	1	0	2	2	1
464	5.12246323	0.07999992	0.10838985	1	1	0	2	2	1
465	3.57088661	0.04905415	0.04905415	2	1	0	2	2	1
466	1.39735794	0.00300000	0.00000000	0	2	0	2	2	1
467	2.88959885	0.04020309	0.04020309	1	2	0	2	2	1
468	1.35715485	0.00000000	0.00000000	2	2	0	2	2	1
469	2.76536179	0.01493454	0.01493454	0	0	1	2	2	1
470	4.23861694	0.09325600	0.11009884	1	0	1	2	2	1
471	2.68704033	0.03177738	0.03177738	2	0	1	2	2	1
472	4.93343830	0.07972240	0.10809708	0	1	1	2	2	1
473	6.30487633	0.25986099	0.41588593	1	1	1	2	2	1
474	4.75329971	0.17924976	0.23574734	2	1	1	2	2	1
475	2.70057392	0.04330921	0.04330921	0	2	1	2	2	1
476	4.07201195	0.20431519	0.24928093	1	2	1	2	2	1
477	2.53956795	0.08827496	0.08827496	2	2	1	2	2	1
478	1.40308762	0.00000000	0.00000000	0	0	2	2	2	1
479	2.87634277	0.02865601	0.02865601	1	0	2	2	2	1
480	1.37443161	0.00300000	0.00000000	2	0	2	2	2	1
481	3.57116413	0.04903889	0.04903889	0	1	2	2	2	1
482	4.94260216	0.17951202	0.23602486	1	1	2	2	2	1
483	3.44069099	0.10555172	0.10555172	2	1	2	2	2	1
484	1.35404873	0.00000000	0.00000000	0	2	2	2	2	1
485	2.72548676	0.08516884	0.08516884	1	2	2	2	2	1
486	1.26887989	0.00000000	0.00000000	2	2	2	2	2	1
487	0.00000000	0.00000000	0.00000000	0	0	0	0	0	2
488	1.55577946	0.00000000	0.00000000	1	0	0	0	0	2
489	0.00300000	0.00000000	0.00000000	2	0	0	0	0	2
490	2.23367310	0.00000000	0.00000000	0	1	0	0	0	2
491	3.76852512	0.02092743	0.02092743	1	1	0	0	0	2
492	2.21274567	0.00000000	0.00000000	2	1	0	0	0	2
493	0.00000000	0.00000000	0.00000000	0	2	0	0	0	2

FILE: SHANON OUT A1 PORTLAND STATE UNIVERSITY VM/SP R4 03/27/87

494	1.53485203	0.00000000	0.00000000	1	2	0	0	0	2
495	0.00000000	0.00000000	0.00000000	2	2	0	0	0	2
496	1.36279774	0.00000000	0.00000000	0	0	1	0	0	2
497	2.86359894	0.04997826	0.04997826	1	0	1	0	0	2
498	1.31281948	0.00000000	0.00000000	2	0	1	0	0	2
499	3.58075428	0.01571655	0.01571655	0	1	1	0	0	2
500	5.04045963	0.11179066	0.13695908	1	1	1	0	0	2
501	3.48468013	0.04088497	0.04088497	2	1	1	0	0	2
502	1.34708118	0.00000000	0.00000000	0	2	1	0	0	2
503	2.80678654	0.07514668	0.07514668	1	2	1	0	0	2
504	1.27193451	0.00000000	0.00000000	2	2	1	0	0	2
505	0.00000000	0.00000000	0.00000000	0	0	2	0	0	2
506	1.50580120	0.00000000	0.00000000	1	0	2	0	0	2
507	0.00000000	0.00000000	0.00000000	2	0	2	0	0	2
508	2.21795654	0.00000000	0.00000000	0	1	2	0	0	2
509	3.67766190	0.04609585	0.04609585	1	1	2	0	0	2
510	2.17186069	0.00000000	0.00000000	2	1	2	0	0	2
511	0.00000000	0.00000000	0.00000000	0	2	2	0	0	2
512	1.45970535	0.00000000	0.00000000	1	2	2	0	0	2
513	0.00000000	0.00000000	0.00000000	2	2	2	0	0	2
514	1.01244926	0.00000000	0.00000000	0	0	0	1	0	2
515	2.55999374	0.00823498	0.00823498	1	0	0	1	0	2
516	1.00421429	0.00000000	0.00000000	2	0	0	1	0	2
517	3.23325253	0.01286983	0.01286983	0	1	0	1	0	2
518	4.73963642	0.06226540	0.08249855	1	1	0	1	0	2
519	3.19385696	0.03310299	0.03310299	2	1	0	1	0	2
520	0.99957943	0.00000000	0.00000000	0	2	0	1	0	2
521	2.50596333	0.02846813	0.02846813	1	2	0	1	0	2
522	0.97111130	0.00000000	0.00000000	2	2	0	1	0	2
523	2.36565971	0.00958729	0.00958729	0	0	1	1	0	2
524	3.85170937	0.07931709	0.09083366	1	0	1	1	0	2
525	2.29592991	0.02110386	0.02110386	2	0	1	1	0	2
526	4.55340576	0.05551434	0.07285500	0	1	1	1	0	2
527	5.95563289	0.20706067	0.31829453	1	1	1	1	0	2
528	4.39985943	0.12992001	0.16474819	2	1	1	1	0	2
529	2.31973267	0.02692795	0.02692795	0	2	1	1	0	2
530	3.72196579	0.15954635	0.18855095	1	2	1	1	0	2
531	2.18711376	0.05593204	0.05593204	2	2	1	1	0	2
532	1.00286198	0.00000000	0.00000000	0	0	2	1	0	2
533	2.48891163	0.01975155	0.01975155	1	0	2	1	0	2
534	0.98311043	0.00000000	0.00000000	2	0	2	1	0	2
535	3.19060302	0.03021049	0.03021049	0	1	2	1	0	2
536	4.59284115	0.13377857	0.17149925	1	1	2	1	0	2
537	3.08703995	0.06793118	0.06793118	2	1	2	1	0	2
538	0.97265148	0.00000000	0.00000000	0	2	2	1	0	2
539	2.37488461	0.05747223	0.05747223	1	2	2	1	0	2
540	0.91517925	0.00000000	0.00000000	2	2	2	1	0	2
541	0.00000000	0.00000000	0.00000000	0	0	0	2	0	2
542	1.54754449	0.00000000	0.00000000	1	0	0	2	0	2
543	0.00000000	0.00000000	0.00000000	2	0	0	2	0	2
544	2.22080326	0.00000000	0.00000000	0	1	0	2	0	2
545	3.72718716	0.04116058	0.04116058	1	1	0	2	0	2
546	2.17964268	0.00000000	0.00000000	2	1	0	2	0	2
547	0.00000000	0.00000000	0.00000000	0	2	0	2	0	2
548	1.50638390	0.00000000	0.00000000	1	2	0	2	0	2

FILE: SHANON OUT A1 PORTLAND STATE UNIVERSITY VM/SP R4 03/27/87

549	0.00000000	0.00000000	0.00000000	2	2	0	2	0	2
550	1.35321045	0.00000000	0.00000000	0	0	1	2	0	2
551	2.83926010	0.06149483	0.06149483	1	0	1	2	0	2
552	1.29171562	0.00000000	0.00000000	2	0	1	2	0	2
553	3.54095650	0.03305721	0.03305721	0	1	1	2	0	2
554	4.94318962	0.17836857	0.22102451	1	1	1	2	0	2
555	3.39564514	0.07571316	0.07571316	2	1	1	2	0	2
556	1.32015324	0.00000000	0.00000000	0	2	1	2	0	2
557	2.72239636	0.10415077	0.10415077	1	2	1	2	0	2
558	1.21600246	0.00000000	0.00000000	2	2	1	2	0	2
559	0.00000000	0.00000000	0.00000000	0	0	2	2	0	2
560	1.48604965	0.00000000	0.00000000	1	0	2	2	0	2
561	0.00000000	0.00000000	0.00000000	2	0	2	2	0	2
562	2.18774605	0.00000000	0.00000000	0	1	2	2	0	2
563	3.58997917	0.08381653	0.08381653	1	1	2	2	0	2
564	2.10392952	0.00000000	0.00000000	2	1	2	2	0	2
565	0.00000000	0.00000000	0.00000000	0	2	2	2	0	2
566	1.40223312	0.00000000	0.00000000	1	2	2	2	0	2
567	0.00000000	0.00000000	0.00000000	2	2	2	2	0	2
568	0.61644459	0.00000000	0.00000000	0	0	0	0	1	2
569	2.16636753	0.00585651	0.00585651	1	0	0	0	1	2
570	0.61058807	0.00000000	0.00000000	2	0	0	0	1	2
571	2.84352303	0.00659466	0.00659466	0	1	0	0	1	2
572	4.36490154	0.04099560	0.04861259	1	1	0	0	1	2
573	2.80912209	0.01421165	0.01421165	2	1	0	0	1	2
574	0.60984993	0.00000000	0.00000000	0	2	0	0	1	2
575	2.13122845	0.01347351	0.01347351	1	2	0	0	1	2
576	0.59637642	0.00000000	0.00000000	2	2	0	0	1	2
577	1.97323513	0.00600719	0.00600719	0	0	1	0	1	2
578	3.46801662	0.06700516	0.07216835	1	0	1	0	1	2
579	1.91223717	0.01117039	0.01117039	2	0	1	0	1	2
580	4.17717934	0.03573608	0.04315376	0	1	1	0	1	2
581	5.60635471	0.16234016	0.23149300	1	1	1	0	1	2
582	4.05057526	0.08557796	0.10488892	2	1	1	0	1	2
583	1.94350624	0.01342487	0.01342487	0	2	1	0	1	2
584	3.37268162	0.11910152	0.13615799	1	2	1	0	1	2
585	1.83782959	0.03048134	0.03048134	2	2	1	0	1	2
586	0.61043739	0.00000000	0.00000000	0	0	2	0	1	2
587	2.10521889	0.01101971	0.01101971	1	0	2	0	1	2
588	0.59941769	0.00000000	0.00000000	2	0	2	0	1	2
589	2.81438160	0.01401234	0.01401234	0	1	2	0	1	2
590	4.24355698	0.09063816	0.11014843	1	1	2	0	1	2
591	2.73775578	0.03352261	0.03352261	2	1	2	0	1	2
592	0.59642506	0.00000000	0.00000000	0	2	2	0	1	2
593	2.02560043	0.03052993	0.03052998	1	2	2	0	1	2
594	0.56589508	0.00000000	0.00000000	2	2	2	0	1	2
595	1.23060799	0.39828587	0.39828587	0	0	0	1	1	2
596	2.77037144	0.41430187	0.41622639	1	0	0	1	1	2
597	1.21459198	0.40021038	0.40021038	2	0	0	1	1	2
598	3.44280815	0.41975880	0.42176723	0	1	0	1	1	2
599	4.93504906	0.48329735	0.51257038	1	1	0	1	1	2
600	3.37926960	0.44827843	0.44903133	2	1	0	1	1	2
601	1.20913506	0.40029430	0.40029430	0	2	0	1	1	2
602	2.70137596	0.44290543	0.44357491	1	2	0	1	1	2
603	1.16652393	0.40096378	0.40096378	2	2	0	1	1	2

FILE: SHANON OUT A1 PORTLAND STATE UNIVERSITY VM/SP R4 03/27/87

604	2.57794762	0.41374397	0.41360760	0	0	1	1	1	2
605	4.05120277	0.49626827	0.51443672	1	0	1	1	1	2
606	2.49542332	0.43219852	0.43191242	2	0	1	1	1	2
607	4.74602413	0.47934055	0.51326847	0	1	1	1	1	2
608	6.11746216	0.66368198	0.82442093	1	1	1	1	1	2
609	4.56168270	0.57369481	0.64007950	2	1	1	1	1	2
610	2.51235104	0.44415951	0.44767199	0	2	1	1	1	2
611	3.88378906	0.60757351	0.65700722	1	2	1	1	1	2
612	2.34893703	0.49048519	0.49359322	2	2	1	1	1	2
613	1.21514988	0.39814949	0.39814949	0	0	2	1	1	2
614	2.68840504	0.43069553	0.43247032	1	0	2	1	1	2
615	1.18260384	0.39992428	0.39992428	2	0	2	1	1	2
616	3.38322639	0.44302952	0.45368671	0	1	2	1	1	2
617	4.75466442	0.53239269	0.64403629	1	1	2	1	1	2
618	3.24886322	0.50552559	0.50967312	2	1	2	1	1	2
619	1.16526985	0.40380669	0.40380669	0	2	2	1	1	2
620	2.53670788	0.49207401	0.49233913	1	2	2	1	1	2
621	1.07700253	0.40407181	0.40407181	2	2	2	1	1	2
622	0.21815872	0.00000000	0.00000000	0	0	0	2	1	2
623	1.75792217	0.00778103	0.00778103	1	0	0	2	1	2
624	0.21037769	0.00000000	0.00000000	2	0	0	2	1	2
625	2.43035889	0.00860310	0.00860310	0	1	0	2	1	2
626	3.92259979	0.06390667	0.07026863	1	1	0	2	1	2
627	2.37505531	0.01496506	0.01496506	2	1	0	2	1	2
628	0.20955563	0.00000000	0.00000000	0	2	0	2	1	2
629	1.70179653	0.01414299	0.01414299	1	2	0	2	1	2
630	0.19541264	0.00000000	0.00000000	2	2	0	2	1	2
631	1.56549935	0.00587082	0.00587082	0	0	1	2	1	2
632	3.03875351	0.08016014	0.08517361	1	0	1	2	1	2
633	1.49120903	0.01088428	0.01088428	2	0	1	2	1	2
634	3.73357487	0.05859756	0.06966400	0	1	1	2	1	2
635	5.10501289	0.23470402	0.32307911	1	1	1	2	1	2
636	3.55746841	0.12426758	0.14697266	2	1	1	2	1	2
637	1.51277161	0.01693726	0.01693726	0	2	1	2	1	2
638	2.88420963	0.15188313	0.16853523	1	2	1	2	1	2
639	1.37782574	0.03358936	0.03358936	2	2	1	2	1	2
640	0.21228790	0.00000000	0.00000000	0	0	2	2	1	2
641	1.68554306	0.01279449	0.01279449	1	0	2	2	1	2
642	0.19949341	0.00000000	0.00000000	2	0	2	2	1	2
643	2.38036442	0.01966953	0.01966953	0	1	2	2	1	2
644	3.75180244	0.13428116	0.15228176	1	1	2	2	1	2
645	2.26575279	0.03767014	0.03767014	2	1	2	2	1	2
646	0.19261837	0.00000000	0.00000000	0	2	2	2	1	2
647	1.56405640	0.03079510	0.03079510	1	2	2	2	1	2
648	0.16182327	0.00000000	0.00000000	2	2	2	2	1	2
649	0.00000000	0.00000000	0.00000000	0	0	0	0	2	2
650	1.54992294	0.00000000	0.00000000	1	0	0	0	2	2
651	0.00000000	0.00000000	0.00000000	2	0	0	0	2	2
652	2.22707844	0.00000000	0.00000000	0	1	0	0	2	2
653	3.74845695	0.02354443	0.02854443	1	1	0	0	2	2
654	2.19853401	0.00000000	0.00000000	2	1	0	0	2	2
655	0.00000000	0.00000000	0.00000000	0	2	0	0	2	2
656	1.52137852	0.00000000	0.00000000	1	2	0	0	2	2
657	0.00000000	0.00000000	0.00000000	2	2	0	0	2	2
658	1.35679054	0.00000000	0.00000000	0	0	1	0	2	2

FILE: SHANNON OUT A1 PORTLAND STATE UNIVERSITY VM/SP R4 03/27/87

659	2.85157204	0.05514145	0.05514145	1	0	1	0	2	2
660	1.30164909	0.00000000	0.00000000	2	0	1	0	2	2
661	3.56073475	0.02313423	0.02313423	0	1	1	0	2	2
662	4.98991013	0.14388180	0.18094349	1	1	1	0	2	2
663	3.43998718	0.06019592	0.06019592	2	1	1	0	2	2
664	1.33365631	0.00000000	0.00000000	0	2	1	0	2	2
665	2.76223169	0.09220314	0.09220314	1	2	1	0	2	2
666	1.24145317	0.00000000	0.00000000	2	2	1	0	2	2
667	0.00000000	0.00000000	0.00000000	0	0	2	0	2	2
668	1.49478149	0.00000000	0.00000000	1	0	2	0	2	2
669	0.00000000	0.00000000	0.00000000	2	0	2	0	2	2
670	2.20394421	0.00000000	0.00000000	0	1	2	0	2	2
671	3.63311958	0.06560612	0.06560612	1	1	2	0	2	2
672	2.13833809	0.00000000	0.00000000	2	1	2	0	2	2
673	0.00000000	0.00000000	0.00000000	0	2	2	0	2	2
674	1.42917536	0.00000000	0.00000000	1	2	2	0	2	2
675	0.00000000	0.00000000	0.00000000	2	2	2	0	2	2
676	0.61416340	0.00000000	0.00000000	0	0	0	1	2	2
677	2.15392685	0.01015949	0.01015949	1	0	0	1	2	2
678	0.60400391	0.00000000	0.00000000	2	0	0	1	2	2
679	2.82636356	0.01467827	0.01467827	0	1	0	1	2	2
680	4.31860447	0.07256031	0.09153843	1	1	0	1	2	2
681	2.76868153	0.03385639	0.03385639	2	1	0	1	2	2
682	0.59928513	0.00000000	0.00000000	0	2	0	1	2	2
683	2.09152603	0.02913761	0.02913761	1	2	0	1	2	2
684	0.57014751	0.00000000	0.00000000	2	2	0	1	2	2
685	1.96150303	0.00945091	0.00945091	0	0	1	1	2	2
686	3.43475819	0.08611870	0.09748554	1	0	1	1	2	2
687	1.88423524	0.02081776	0.02081776	2	0	1	1	2	2
688	4.12957954	0.06845284	0.08944225	0	1	1	1	2	2
689	5.50101757	0.24693775	0.36979961	1	1	1	1	2	2
690	3.95109463	0.15309238	0.19131470	2	1	1	1	2	2
691	1.90250111	0.03044033	0.03044033	0	2	1	1	2	2
692	3.27393913	0.18038082	0.20898056	1	2	1	1	2	2
693	1.75256062	0.05904007	0.05904007	2	2	1	1	2	2
694	0.60471249	0.00000000	0.00000000	0	0	2	1	2	2
695	2.07796764	0.02152634	0.02152634	1	0	2	1	2	2
696	0.58318615	0.00000000	0.00000000	2	0	2	1	2	2
697	2.77278900	0.03586769	0.03586769	0	1	2	1	2	2
698	4.14422703	0.15921116	0.19542217	1	1	2	1	2	2
699	2.64944553	0.07207870	0.07207870	2	1	2	1	2	2
700	0.56684480	0.00000000	0.00000000	0	2	2	1	2	2
701	1.94028282	0.05773735	0.05773735	1	2	2	1	2	2
702	0.51110744	0.00000000	0.00000000	2	2	2	1	2	2
703	0.00000000	0.00000000	0.00000000	0	0	0	2	2	2
704	1.53976345	0.00000000	0.00000000	1	0	0	2	2	2
705	0.00000000	0.00000000	0.00000000	2	0	0	2	2	2
706	2.21220016	0.00000000	0.00000000	0	1	0	2	2	2
707	2.70444107	0.04752254	0.04752254	1	1	0	2	2	2
708	2.16467762	0.00000000	0.00000000	2	1	0	2	2	2
709	0.00000000	0.00000000	0.00000000	0	2	0	2	2	2
710	1.49224091	0.00000000	0.00000000	1	2	0	2	2	2
711	0.00000000	0.00000000	0.00000000	2	2	0	2	2	2
712	1.34733963	0.00000000	0.00000000	0	0	1	2	2	2
713	2.82059479	0.06650829	0.06650829	1	0	1	2	2	2

FILE: SHANON CUI A1 PORTLAND STATE UNIVERSITY VM/SP R4 03/27/87

714	1.26083134	0.00000000	0.00000000	2	0	1	2	2	2
715	3.51541615	0.04412365	0.04412365	0	1	1	2	2	2
716	4.88685417	0.21244907	0.26674366	1	1	1	2	2	2
717	3.34709072	0.09841824	0.09841824	2	1	1	2	2	2
718	1.30321593	0.00000000	0.00000000	0	2	1	2	2	2
719	2.67465401	0.12080288	0.12080288	1	2	1	2	2	2
720	1.18241310	0.00000000	0.00000000	2	2	1	2	2	2
721	0.00000000	0.00000000	0.00000000	0	0	2	2	2	2
722	1.47325516	0.00000000	0.00000000	1	0	2	2	2	2
723	0.00000000	0.00000000	0.00000000	2	0	2	2	2	2
724	2.16807652	0.00000000	0.00000000	0	1	2	2	2	2
725	3.53951454	0.10181713	0.10181713	1	1	2	2	2	2
726	2.06625938	0.00000000	0.00000000	2	1	2	2	2	2
727	0.00000000	0.00000000	0.00000000	0	2	2	2	2	2
728	1.37143803	0.00000000	0.00000000	1	2	2	2	2	2
729	0.00000000	0.00000000	0.00000000	2	2	2	2	2	2