1971

Prepurchase Information Gathering By Household Durable Buyers: An Exploratory Study Using Numerical Taxonomic Analysis

John David Claxton

Follow this and additional works at: https://ir.lib.uwo.ca/digitizedtheses

Recommended Citation
https://ir.lib.uwo.ca/digitizedtheses/519

This Dissertation is brought to you for free and open access by the Digitized Special Collections at Scholarship@Western. It has been accepted for inclusion in Digitized Theses by an authorized administrator of Scholarship@Western. For more information, please contact tadam@uwo.ca, wlswadmin@uwo.ca.
The author of this thesis has granted The University of Western Ontario a non-exclusive license to reproduce and distribute copies of this thesis to users of Western Libraries. Copyright remains with the author.

Electronic theses and dissertations available in The University of Western Ontario’s institutional repository (Scholarship@Western) are solely for the purpose of private study and research. They may not be copied or reproduced, except as permitted by copyright laws, without written authority of the copyright owner. Any commercial use or publication is strictly prohibited.

The original copyright license attesting to these terms and signed by the author of this thesis may be found in the original print version of the thesis, held by Western Libraries.

The thesis approval page signed by the examining committee may also be found in the original print version of the thesis held in Western Libraries.

Please contact Western Libraries for further information:
E-mail: libadmin@uwo.ca
Telephone: (519) 661-2111 Ext. 84796
Web site: http://www.lib.uwo.ca/
PREPURCHASE INFORMATION GATHERING
BY HOUSEHOLD DURABLE BUYERS
AN EXPLORATORY STUDY USING NUMERICAL TAXONOMIC ANALYSIS

by
John David Claxton
School of Business Administration

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Canada
June 1971

© John David Claxton 1971
ABSTRACT

Two of the considerations confronting a manager as he plans marketing communications are - what information is of interest to buyers, and how do buyers obtain this information. Answers to these questions can be used to tailor communications to match buyer interests. This research studies buyer information gathering in terms of types and sources of information considered and time spent on the information gathering process. The general purpose was to describe this process by identifying groups of buyers with similar information gathering styles.

A sample of 546 housewives from one metropolitan area was interviewed with respect to their prepurchase information gathering. Each had recently purchased a major appliance or piece of furniture. Numerical taxonomic procedures were used to identify buyer groups with similar information gathering styles.

The research was considered exploratory for two reasons. The first was the somewhat arbitrary nature of numerical taxonomic procedures. The second was the relatively untested nature of buyer groups based on information gathering styles. The objective was to
make progress in three areas. The first was the identification of major information gathering styles and key factors differentiating these styles. The second was the measurement of information gathering to facilitate style identification. The third area was the evaluation and development of numerical taxonomic procedures.

The findings indicated three factors were useful in differentiating information gathering styles - total store visits, time spent, and number of sources considered. Several major styles were identified. For example a "thorough" group of buyers made eight store visits, considered the purchase for months, and used three sources. A "non-thorough" group made one visit, considered the purchase for days, and used only one source. Consideration of the findings suggested other measurement areas of interest in differentiating information gathering styles.

Numerical taxonomic procedures provided a relatively unstructured approach to the identification of "natural" information gathering groupings. However, considerable research effort was required to select appropriate procedures from the range available, and to evaluate the meaning of groups identified. As part of this research two procedures were developed to facilitate evaluation of groups. Reverse matrix
clustering proved useful in directing attention to concentrations in the subject-scatter. Overlap assessment provided a visual guide to the distinctiveness of groups and to the extent groups replicated across samples.
ACKNOWLEDGMENT

A dissertation is a summation of energies and resources. I am pleased to acknowledge contributions that led to the completion of this one. Support from the Samuel Bronfman Foundation and the Associates' Research Fund made possible the survey of household durable buyers. Professors B. Portis and K.G. Hardy provided critical assessment both when the research was being proposed and as the dissertation was being completed. I particularly appreciate the support of Professor J.N. Fry who supervised this dissertation. His pointed queries directed attention to areas that otherwise might have been neglected.

I also appreciate the efforts of Mrs. Gwen Pemberton and Mrs. Marilyn Ambrose. Mrs. Pemberton provided needed editorial review. Mrs. Ambrose applied great care and attention to the typing of the final dissertation.

Finally, I thank my wife, Dawna. She edited, typed and proofread many drafts, and above all provided continuous morale support.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Certificate of Examination</td>
<td>ii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENT</td>
<td>vi</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiii</td>
</tr>
<tr>
<td>Chapter 1 - INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Prepurchase Information Gathering</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Outline of Research</td>
<td>3</td>
</tr>
<tr>
<td>1.21 Subject Area</td>
<td>3</td>
</tr>
<tr>
<td>1.22 Approach</td>
<td>4</td>
</tr>
<tr>
<td>1.23 Analysis and Findings</td>
<td>5</td>
</tr>
<tr>
<td>Chapter 2 - PREPURCHASE INFORMATION GATHERING</td>
<td>7</td>
</tr>
<tr>
<td>2.1 Background</td>
<td>7</td>
</tr>
<tr>
<td>2.11 Conceptual Model of Prepurchase Information Gathering</td>
<td>8</td>
</tr>
<tr>
<td>2.12 The Limited Nature of Prepurchase Information Gathering</td>
<td>10</td>
</tr>
<tr>
<td>2.13 Prepurchase Information Gathering Styles</td>
<td>14</td>
</tr>
<tr>
<td>2.14 Background: Summary</td>
<td>20</td>
</tr>
<tr>
<td>2.2 Modeling Prepurchase Information Gathering</td>
<td>21</td>
</tr>
<tr>
<td>2.21 Level of Measurement Aggregation</td>
<td>21</td>
</tr>
</tbody>
</table>
5.3 Evaluation of Variables: Features. 68
5.4 Evaluation of Variables: Store Visits. 71
5.5 Evaluation of Variables: Time. 75
5.6 Preliminary Measurement Analysis: Summary. 75

Chapter 6 - NUMERICAL TAXONOMIC PROCEDURES:
PRELIMINARY ANALYSIS. 79
6.1 Proximity Measure Selection. 79
6.2 Clustering Algorithm Selection. 83
6.3 Evaluation of Clustering Results. 85
   6.3.1 Clustering a Proximity Matrix Containing Ties. 86
   6.3.2 Procedures for Assessing "Appropriate" Number of Clusters. 92
   6.3.3 Overlap Assessment. 97
6.4 Summary. 109

Chapter 7 - NUMERICAL TAXONOMIC ANALYSIS OF PRE-
PURCHASE INFORMATION GATHERING. 111
7.1 Standard Analysis Procedures. 111
7.2 Exploration Analysis: Nominal Variables. 113
   7.2.1 18-Variables. 117
   7.2.2 14-Variables. 123
   7.2.3 12-Variables. 128
   7.2.4 9-Variables. 135
   7.2.5 Nominal Variable Approach: Summary. 139
7.3 Exploration Analysis: Interval Variables. 141
   7.3.1 5-Variables. 144
   7.3.2 4-Variables. 150
LIST OF TABLES

Table | Page
-----|-----
5.1 Measurement of Alternatives: Correlation Matrix. | 63
5.2 Measurement of Information Sources: Phi Correlation Matrix. | 67
5.3 Measurement of Features | 69
5.4 Measurement of Features: Phi Correlation Matrix. | 70
5.5 Measurement of Store Visits | 72
5.6 Measurement of Store Visits: Product Moment Correlation Matrix. | 73
5.7 Measurement of Time | 76
7.1 Significant Phi Correlations: 18 Nominal Variables. | 114
7.2 Subgroup Comparisons: 18 Nominal Variables. | 116
7.3 Exploration Analysis: Nominal Approaches. | 118
7.4 Product Moment Correlation: 5 Interval Variables. | 142
7.5 Subgroup Comparisons: 5 Interval Variables. | 143
7.6 Exploration Analysis: Interval Approaches. | 145
7.7 Furniture and Appliance Sample Means. | 170
7.8 Product Moment Correlation: 3 Interval Variables: Appliance Data | 173
7.9 Subgroup Comparisons: 3 Interval Variables: Appliance Data. | 174
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Example of Clustering using Johnson's Diameter Algorithm</td>
<td>41</td>
</tr>
<tr>
<td>4.1</td>
<td>Questions Measuring Information Gathering</td>
<td>54</td>
</tr>
<tr>
<td>4.2</td>
<td>Research Design</td>
<td>56</td>
</tr>
<tr>
<td>5.1</td>
<td>Measurement of Alternatives</td>
<td>62</td>
</tr>
<tr>
<td>5.2</td>
<td>Measurement of Information Sources</td>
<td>66</td>
</tr>
<tr>
<td>5.3</td>
<td>Information Gathering Measurement: Final Variable Set</td>
<td>78</td>
</tr>
<tr>
<td>6.1</td>
<td>Reverse Matrix Clustering</td>
<td>90</td>
</tr>
<tr>
<td>6.2</td>
<td>Overlap Assessment Procedure: Example One</td>
<td>100</td>
</tr>
<tr>
<td>6.3</td>
<td>Overlap Assessment Procedure: Example Two</td>
<td>106</td>
</tr>
<tr>
<td>7.1</td>
<td>Clustering Based on 18 Nominal Variables: Group A</td>
<td>120</td>
</tr>
<tr>
<td>7.2</td>
<td>Clustering Based on 18 Nominal Variables: Group B</td>
<td>121</td>
</tr>
<tr>
<td>7.3</td>
<td>Overlap Assessment: 18 Nominal Variables</td>
<td>122</td>
</tr>
<tr>
<td>7.4</td>
<td>Clustering Based on 14 Nominal Variables: Group A</td>
<td>124</td>
</tr>
<tr>
<td>7.5</td>
<td>Clustering Based on 14 Nominal Variables: Group B</td>
<td>125</td>
</tr>
<tr>
<td>7.6</td>
<td>Overlap Assessment: 14 Nominal Variables</td>
<td>126</td>
</tr>
<tr>
<td>7.7</td>
<td>Clustering Based on 12 Nominal Variables: Group A</td>
<td>129</td>
</tr>
<tr>
<td>7.8</td>
<td>Clustering Based on 12 Nominal Variables: Group B</td>
<td>130</td>
</tr>
<tr>
<td>7.9</td>
<td>Overlap Assessment: 12 Nominal Variables</td>
<td>131</td>
</tr>
</tbody>
</table>
7.10 Clustering Based on 9 Nominal Variables: Group A. .................................. 136
7.11 Clustering Based on 9 Nominal Variables: Group B. .................................. 137
7.12 Overlap Assessment: 9 Nominal Variables. .............................................. 138
7.13 Clustering Based on 5 Interval Variables: Group A. ................................. 147
7.14 Clustering Based on 5 Interval Variables: Group B. ................................. 148
7.15 Overlap Assessment: 5 Interval Variables .............................................. 149
7.16 Clustering Based on 4 Interval Variables: Group A. ................................. 151
7.17 Clustering Based on 4 Interval Variables: Group B. ................................. 152
7.18 Overlap Assessment: 4 Interval Variables .............................................. 153
7.19 Clustering Based on 3 Interval Variables: Group A. ................................. 157
7.20 Clustering Based on 3 Interval Variables: Group B. ................................. 158
7.21 Overlap Assessment: 3 Interval Variables: Distinctiveness. ...................... 159
7.22 Overlap Assessment: 3 Interval Variables: Replication. .......................... 160
7.23 Overlap Assessment: 3 Interval Variables: 3-Dimensional Representation ...... 161
7.24 Clustering Based on 3 Interval Variables: Group A: Re-test ...................... 176
7.25 Clustering Based on 3 Interval Variables: Group B: Re-test ...................... 177
7.26 Overlap Assessment: 3 Interval Variables: Appliance Data Distinctiveness ... 178
7.27 Overlap Assessment: 3 Interval Variables: Appliance Data Replication ....... 179
CHAPTER 1
INTRODUCTION

1.1 PREPURCHASE INFORMATION GATHERING

A buyer considering the purchase of a household durable, such as a living room suite or a refrigerator, is faced with a relatively complex problem. Because the price is several hundred dollars, the buyer probably considers the purchase to be an important transaction. Further, since this type of product is purchased infrequently she (or he) probably has limited shopping experience. Given the importance of the purchase and the lack of experience, how does she decide what to buy? This decision involves two areas of complexity. First, the buyer is faced with many product differences, for example materials, features, styling, color, brand, size, and price. Second, a variety of sources of information are available. Potentially she can consider sources such as advertisements, catalogues, family discussion, retail salesmen, Consumer Reports, discussion with friends, and retail displays.

Given this complexity, buyers cannot consider every possible type of information from every possible source. Thus, the question of concern to marketing managers is: what are the major ways in which information gathering is
accomplished? In particular, is it possible to identify major information gathering processes, and use them as a basis for marketing strategy? For example, suppose two major information gathering processes were identified. One group of buyers, group A, was interested in styling and special features; group B was interested in durability and price. Group A obtained information from magazine advertisements and retail displays; group B, from salesmen and newspapers. The marketing strategy designed to attract group A would emphasize style and features. Major portions of marketing expenditures would be allocated to magazine advertisements, dealer displays and point-of-purchase material. On the other hand the strategy designed to attract group B would emphasize durability and competitive price. Major portions of marketing expenditures would be allocated to newspaper advertisements and retail salesforce development.

Although this is a very simple example, it indicates the potential value of studying buyer information gathering. Marketing managers have recognized this potential; for example an article in Business Week quotes the manager of the Houseware Division of General Electric:

"GE is focusing on how today's consumer makes specific buying decision....where she gets her information....what magazines and newspapers does the consumer read....how much shopping-around does she do?"¹

¹Business Week, April 24, 1971, page 56.
It is also useful to consider briefly marketing literature with respect to information gathering. For the most part buyer behavior theory has concentrated on the description and explanation of what is purchased; there has been relatively little consideration of how the purchase process is carried out. For example, The Theory of Buyer Behavior, developed by Howard and Sheth (1969), suggested several factors that led to "overt search", but did not describe the overt search process. A major exception to the lack of emphasis on the "how" aspects of search was empirical work done by Katona and Mueller (1955). They surveyed recent buyers of household durables to gain an understanding of the nature of alternative processes of decision making. Their research made an important contribution to the description of pre-purchase information gathering.

1.2 OUTLINE OF RESEARCH

1.2.1 Subject Area

The general subject area considered in this research was prepurchase information gathering by household durable buyers. The central purpose was to work toward a descriptive model of the information gathering process. In particular the research effort focused on two questions. First, was it possible to categorize buyers based on differences in their information gathering processes? Second, what were the key factors that distinguished one
process from another? Prepurchase information gathering is defined in Chapter 2. Also in this chapter a conceptual model is developed, and related literature is reviewed.

1.22 Approach

Numerical taxonomy was the analytical approach selected for modeling prepurchase information gathering. The objective of this approach was to identify buyer groups so that within each group the information gathering processes were similar, and between groups the processes were different. The application of numerical taxonomic procedures involved three steps: the development of an inter-subject proximity matrix; the selection and application of a clustering algorithm; and the evaluation of the clustering results. A description of these steps is presented in Chapter 3.

The research was considered exploratory for two reasons. First, categorizing buyers based on differences in information gathering was relatively untested. Second, there were weaknesses in numerical taxonomic procedures that had to be addressed. Thus, the research did not set out to test a series of hypotheses. The intention was to explore three major areas: measurement of prepurchase information gathering to facilitate identification of buyer groups, assessment of numerical taxonomic procedures, and taxonomic modeling of information gathering done by
household durable buyers. The exploratory nature of the research is described in Chapter 4. The data base and sequence of analysis is also discussed in this chapter, and a flow chart of the research is presented.

1.23 Analysis and Findings

The areas of analysis correspond to the three major areas mentioned above. The analysis of survey measures of prepurchase information gathering is described in Chapter 5. This analysis led to the selection of a set of variables to be used as a base for taxonomic modeling. The assessment of numerical taxonomic procedures is presented in Chapter 6. Consideration of available procedures, plus development of new ones resulted in the final set used in this research. The final area of analysis, numerical taxonomic modeling of prepurchase information gathering, is discussed in Chapter 7. This chapter describes two sections of analysis, the exploration analysis and the re-test analysis. The exploration analysis attempted to identify buyer groups based on several different subsets of the information gathering variables. The re-test analysis repeated the most fruitful parts of the exploration analysis using a different sample.

The final chapter presents conclusions and implications with respect to the three areas on analysis. First, progress made toward a model of prepurchase information
gathering is considered in terms of future empirical research in this area, and as an approach to market segmentation. Second, measurement of information gathering is reviewed in terms of the relative strengths of the measures used, and implications for measurement in future research. Third, discussion of numerical taxonomic analysis outlines the merits of these procedures for modeling information gathering, and for other marketing applications.
CHAPTER 2

THE PREPURCHASE INFORMATION GATHERING PROCESS

The central purpose of the research was to work toward a descriptive model of prepurchase information gathering. This chapter provides a framework for the research by presenting a general description of the information gathering process, and discussing methods for modeling it.

2.1 BACKGROUND

The prepurchase information gathering process was defined as the set of activities directed toward gathering information to facilitate a purchase decision. Alternative terms used in this dissertation are information gathering process, information gathering, prepurchase search process, search process, and search. The limits of the process were from the time a buyer started thinking and talking about buying until she made the purchase. The process included activities that ranged from noticing an advertisement while reading the paper to seeking out every sales outlet in town.
There were three concepts important to description of the information gathering process. First, information gathering was seen as multidimensional. Its description required identification of such aspects as: what information was gathered, where was the information obtained, and what was the duration of search. Second, information gathering was seen as limited. For example, buyers did not use all of the available information sources, nor gather information on all possible product features. The third concept considered important was that of information gathering styles. The styles concept introduced the possibility of identifying buyer groups based on particular combinations of information gathering activities. These three concepts are discussed in the next three subsections of this chapter.

2.11 Conceptual Model of Prepurchase Information Gathering

Prior to the discussion of the information gathering model four terms require definition. Information gathering is described in terms of dimensions, characteristics, factors, and areas. Dimension is used when the aspect being discussed is continuous. For example, "number of store visits" would be relatively continuous. Characteristic is used when the aspect being discussed contains categories. "Information from friends" with yes-no response categories would be an example. Factor
is used to refer to either dimensions or characteristics. Area is used as a general term that may include several information gathering factors.

A simple model of the information gathering process was developed to provide a focus for the research. The model contained three primary areas: type, source, and time. Type referred to the content of the information gathered, for example, price information, style information, quality information. Source referred to where the information was obtained, for example, newspapers, friends and salesmen. Time referred to the period over which information was gathered.

The model also contained two secondary areas. These were considered to be a function of the primary areas. That is, total amount was a summation of the quantity of each type that was obtained from each source. Intensity was the total amount divided by time. The distinction between primary and secondary was included to emphasize areas of measurement. The former represented areas that measurement attempted to tap. The secondary areas were included to provide a more complete picture of the information gathering process.

In summary, the central interest was a description of information gathering that included: what information was gathered (type), where the information was obtained (source), and how long the process lasted (time).
Comparison of this conceptual model and buyer behavior models in the marketing literature indicated a considerable difference in emphasis. Howard and Sheth (1969) in *The Theory of Buyer Behavior* suggested factors that they felt resulted in "overt search". Andreasen's consumer decision model described factors influencing the amount of information gathered (Andreasen, 1968). Cox (1967) considered information gathering to be a method of reducing "Perceived Risk". All of these models discussed factors that were considered to lead to search, but stopped short of a description of the search process. The information gathering model presented above emphasized this description.

2.12 The Limited Nature of Prepurchase Information Gathering

Buyers cannot consider all available types of product information nor all possible sources. Thus, the
second concept considered in the description of information gathering was its limited nature. Writings from economics, psychology and marketing helped to describe and explain the limited nature of search. In economic theory the cost of search was suggested as a reason for limiting information gathering. Stigler (1961) discussed the economics of information and prescribed that search should be continued only as long as the marginal return expected from further search was greater than the marginal cost. Mincer (1964) also recognized the importance of search costs. He suggested that in the economic analysis of individual consuming units the resources expended on search (particularly shopper's time) should be included in the price of goods.

Limited cognitive capacity also provided an explanation for the limited nature of information gathering. Bruner, Goodnow and Austin (1956) discussed limited cognitive capacity in terms of limited span of attention and limited immediate memory. It seemed intuitively that either of these factors could result in limiting of the information gathering process. Evoked set, a concept similar to limited immediate memory, was used by March and Simon (1958) in their description of decision making. They identified the evoked set as the part of memory that exerted significant influence on behavior at a point in time. Because of the limited nature of the evoked set,
they considered the decision maker to have a limited awareness of the environment, and a limited knowledge of alternatives and results.

Cost of search and limited cognitive capacity were considered to be potential causes of limited prepurchase activity. There were also several theoretical underpinnings for the manner in which prepurchase information gathering was limited. Cyert and March (1963) discussed ways in which decision making was simplified in their description of decision making in a firm. They suggested that goals are considered sequentially, thus decisions are simplified by removing the possibility of goal-conflict. An analogy in buyer decision making would be the sequential consideration of purchase decisions. For example, a vacation may not be seen as an alternative to a new refrigerator, although limited resources may preclude both expenditures. Sequential purchases would simplify the search process by removing the need to look for information comparing alternative product types.

Bauer (1963) discussed limited search in terms of choice of information sources. He suggested that the audience evaluated the communicator, and then selected the communication to which they would attend on the basis of this evaluation.

Bruner, Goodnow and Austin (1956) discussed limited search in terms of selection of attributes for product
comparison. They described the search for attributes as "cue searching", suggesting that the cues selected were a function of the predictive value of the cue, and the salience of the cue to the decision maker. Cox (1967) extended the cue search concept in his "Sorting Rule Model". The purpose of his model was to explain how consumers selected the cues they used in making purchase decisions.

The limited nature concept has also been used in models of buyer behavior. The Howard and Sheth model (1969) suggested the use of a limited number of "choice criteria" when selecting one brand from among several. The model developed by Engle, Kollat and Blackwell (1968) indicated that limited response sets influence the search for alternatives, evaluation of alternatives, and purchase process.

The writings reviewed provided a theoretical description, and explanation of decision making as a limited activity. Survey data also supported the "limited nature" concept. Katona and Mueller (1955, page 30) based on a sample of 360 households from all across the United States, found a variety of ways buyers of major household durables limited their prepurchase search. Of the families sampled 70 percent did not consider saving nor some other product as an alternative to the purchase under consideration. This supported the "limited alternatives" suggested by
Cyert and March (1963). With respect to limited sources of information suggested by Bauer (1963), only 33 percent of the sample reported considering more than one source of information, and only 40 percent visited more than one store. Finally, 57 percent of the sample considered only one price range, and only 35 percent considered more than one product feature; this supported the concept of limited product cues.

Using a telephone sample of 900 from one city, Dommermuth (1965) gathered data on the recent purchase of household appliances. His findings were similar to those of Katona and Mueller. He found that only 45 percent of the sample visited more than one store, and only 44 percent reported considering more than one brand.

Review of theoretical writings and empirical research provided evidence supporting the concept of limited prepurchase information gathering. The potential causes of limited search were cost of search and limited cognitive capacities. Three ways of limiting search also were identified; they were limiting product alternative, information sources, and product cues.

2.13 Prepurchase Information Gathering Styles

Prepurchase information gathering was seen as multidimensional and limited. The third concept, information gathering styles, was added to allow for differences
across buyers in the extent to which dimensions were limited. An information gathering style was defined as a particular information gathering process, that is, a particular combination of types, sources and time. For example, if these three areas were measured as dichotomous variables (many-few types, many-few sources, and long-short time), eight information gathering styles would be possible. Several terms are used interchangeably in this dissertation: prepurchase information gathering styles, information gathering styles, prepurchase search styles, search styles, and styles.

There were two theoretical explanations for the existence of prepurchase styles. The first explained styles in terms of individual differences; the second, in terms of situational differences. Kelman and Cohler (1959) discussed individual differences in terms of differences in cognitive structure. They considered cognitive structure to include cognitive needs and cognitive style. In particular an individual's effort to reduce uncertainty was considered to be a function of his need for cognitive clarity. An individual's approach to reducing uncertainty was considered a function of his cognitive style - for example "clarifiers" would attempt to gain more information and "simplifiers" would attempt to simplify the environment. Thus, differences in need for cognitive clarity would be expected to influence
search style by influencing the amount of information gathered. Cognitive style would be expected to influence the range of information types gathered and information sources used. Clarifiers would handle many types and sources, and simplifiers would handle few.

The second concept that was useful in the explanation of individual differences in search styles was the evoked set concept used by March and Simon (1958). Based on this concept an individual would consider a limited set of information types, and a limited set of sources. Various information gathering styles would be expected if the evoked set of types and sources were assumed to be stable over the pre-purchase period, and if individual differences were assumed in evoked sets. Support for the stability assumption was discussed by Newell, Shaw, and Simon (1958). They considered the evoked set to change very little during the course of a decision making process. Support for the assumption of individual differences in evoked sets could be hypothesized on the basis of interests and abilities. For example an automobile buyer interested in economy and sportiness would probably have a different evoked product set from a buyer interested in luxury and prestige. Further, the evoked set of information sources of a buyer with considerable technical ability would probably be different from the evoked set of a buyer
without these skills. Although it seems reasonable to assume the stability of the evoked set, and individual differences in evoked sets, there appears to be no empirical evidence supporting these assumptions.

The final source of individual differences explaining information gathering styles, was differences in product knowledge. Howard (1963) used this notion as the basis for dividing purchase decision making into three categories, extensive problem solving, limited problem solving, and automatic response. Extensive problem solving was described as the situation where the buyer had neither an established set of choice criteria, nor an established set of alternatives. Limited problem solving was the situation where the buyer had both choice criteria and a set of alternatives. Finally, automatic response was the situation where the buyer had rated his brand alternatives on his choice criteria. Information gathering style was expected to be different for each of these three situations.

The second potential cause of differences in information gathering styles was termed situational factors. The first factor in this category was immediacy of need. This was expected to influence the time spent on information gathering and possibly the amount of information gathered. For example, a buyer with a broken refrigerator was probably in a greater
hurry to get a new one than was a buyer with a working model. The second situational factor that potentially resulted in differences in prepurchase styles was economic constraints. It was likely this factor would influence the type of information gathered (particularly price and credit information), and the price range of the alternatives considered.

In addition to the theoretical support provided by the concepts discussed, several marketing research studies supported the concept of information gathering styles. Convincing evidence of styles was work done by Dommermuth (1965) on "The Shopping Matrix". He compared prepurchase behavior using a matrix which had as one dimension number of stores visited, and as the other number of brands considered. His data displayed a considerable range of prepurchase styles, for example, considered one brand at one store, five or more brands at five or more stores, four brands at one store, and one brand at three stores.

Another market study supporting the style concept was Katona and Mueller's (1955) study of buyers of household durables. After analysing several indices of "deliberation" Mueller concluded that there appeared to be four prepurchase styles, thorough attention to all aspects of deliberation, moderate attention to all aspects, limited attention to all aspects, and a style
category she referred to as "feature substitution". In the latter category there was thorough attention to some aspects of deliberation and limited attention to others.

Newman (1969) also provided evidence that supported the buyer styles concept. Based on survey data describing the purchase of major durables, he suggested that the variable describing duration of search appeared to be bimodal. This indicated the possibility of two very general prepurchase styles.

Finally, research done by Cox (1967) showed that at the individual level there were definite patterns or styles of prepurchase behavior. He did extensive interviews with two subjects and found for example, the pattern used by one subject was to consult Consumer Reports and then shop at discount stores for name brand products.

This subsection discussed the information gathering style concept. The concept was based on two general factors, individual differences, and situational differences. Individual differences suggested styles to be a function of cognitive structure, evoked set, and product knowledge. Situational factors resulting in differences in search styles were immediacy of need, and economic constraints. Dommermuth's "Shopping Matrix", Mueller's "feature substitution", Newman's bimodal
"duration", and Cox's "individual decision rules" provided research evidence that also supported the information gathering styles concept.

2.14 Background: Summary

The central area of study was the prepurchase information gathering process. Three concepts were used to describe this process. First, information gathering was described as multidimensional, and a general model of information gathering was developed. The model identified types of information gathered, sources of information used, and time spent gathering as three areas of primary importance for research measurement. Second, information gathering was described as limited. A buyer could consider neither all types, nor all sources of information. Finally, the concept of information gathering styles was discussed. The style concept was introduced to draw attention to differences across buyers in the manner that information gathering was limited. For example, one buyer may have considered a wide range of information types, considered only one source, and gathered information over an extended period of time; another may have considered only one information type, used many sources, and gathered information over a limited period. These would represent two possible information gathering styles, in other words two
particular combinations of types, sources and time.

2.2 MODELING PREPURCHASE INFORMATION GATHERING

A central question was, how can a set of variables measuring information types, sources and time be aggregated, simplified, summarized, etc., to facilitate understanding of the information gathering process? This section discusses two major considerations in modeling this process. The first is the level of measurement aggregation, for example single variables or indices based on the addition of several variables. The second consideration discussed is the level of sample aggregation, for example emphasis on individuals or groups. The final subsection introduces numerical taxonomy as a means of modeling prepurchase information gathering.

Throughout the discussion a distinction is made between descriptor and predictor variables. Descriptor variable refers to measures of types, sources, and time. Predictor variable refers to other measures, such as product knowledge and demographics.

2.2.1 Level of Measurement Aggregation

Variables describing prepurchase information gathering can be treated in three major ways: as single variables, as summed indices, and as profiles. The single variable approach refers to models that describe information gathering using one descriptor variable at a
time. Newman and Staelin (1971) used this approach in their research on factors related to "duration". They used two techniques, automatic interaction detection (AID), and multiple classification analysis (MCA) to determine the predictor variables that were related to duration of the purchase decision.

The summed index approach assumes that there is one dimension underlying a set of search variables. Thus, by combining the set in some fashion (such as simple addition) a general index of information gathering is obtained. Katona and Mueller (1955) used this approach to develop their summary variable, deliberation.

Finally, a set of descriptor variables can be treated as a profile. A profile model simultaneously considers a subject's level on each of the variables in the set. Using this approach Domermuth (1965) categorized subjects based on "number of stores visited", and "number of brands considered". An advantage of a profile model is more complete use of the information provided by the descriptors. For example, treating Domermuth's variables as dichotomous (0-1), a single variable model would result in two subject categories (0 and 1), a summed model would result in three categories (0, 1, and 2), and a profile model would result in four categories (0-0, 0-1, 1-0, and 1-1).
2.22 Level of Sample Aggregation

The second dimension useful in differentiating prepurchase search models is the level of sample aggregation. The model may attempt to describe the total sample, various subsample groups, or individuals. The mean, mode, variance, and range of the total sample on a particular descriptor variable are examples of models describing the total sample. At the other end of the spectrum a multiple regression model, explaining a descriptor variable in terms of a number of predictor variables, is a model where the focus is on individual buyers. Between these two extremes are models where the emphasis is on groups. An example of a group model would be the use of multiple discriminant analysis to identify socio-economic differences between one group of buyers that gathered information over an extended period, and another group that gathered information over a limited period.

One further aspect of group models needs clarification. They fall into two categories with respect to objective. One objective is to identify groups, the other is to describe groups. The example discussed above, discriminant analysis to distinguish between extended and limited search groups, is an example of a descriptive model. In this case the groups are determined prior to the analysis, and the purpose is to
describe the groups in terms of other variables. When the purpose is to identify groups, the general nature of the modeling process is to compare subjects across descriptor variable(s), attempting to identify groups that are similar. For example, when groups are to be based on one descriptor variable, the modes of a histogram could be used to identify groups. When two descriptor variables are being considered, a scatter diagram could be used to identify groups. When more than two descriptor variables are being considered, models more elaborate than histograms or scatter diagrams are required. Numerical taxonomy addresses this situation. The next subsection describes numerical taxonomy and discusses the use of these procedures in the description of prepurchase information gathering.

2.23 Numerical Taxonomic Modeling Of Prepurchase Information Gathering

The general purpose of numerical taxonomic procedures is to identify "natural" groups (or classes or clusters) of subjects. Subjects are compared in terms of a set of descriptor variables. The result of the comparison is the identification of subject groups so that subjects in any particular group are similar. Inverse factor analysis or Q-factoring is a relatively common example of numerical taxonomic analysis. Factor analysis is usually applied to identify major dimensions
that underlie in a set of variables; inverse factor analysis attempts to identify major types that underlie a set of subjects.

Numerical taxonomic analysis involves three steps. The first is to compare every subject pair across a number of descriptor variables, and thereby develop a measure of "likeness" for each pair. The second step is to analyse these likeness measures in order to identify clusters of "like" subjects. The final step is to assess the meaning of the clusters identified. Referring back to the classification of models discussed earlier, this approach to modeling would be classified as a group/profile model. That is, the model emphasizes the comparison of subject groups, and the group comparisons are made in terms of descriptor variable profiles. Chapter 3 presents a more complete description of numerical taxonomic procedures.

In the application of numerical taxonomy to pre-purchase search, the measure of "likeness" would be based on comparison across a set of variables dealing with types, sources and time. Each buyer group identified would represent a particular information gathering style.

2.24 Modeling Prepurchase Information Gathering: Summary

Modeling prepurchase information gathering was discussed with respect to level of measurement aggre-
gation, and level of sample aggregation. The discussion of measurement presented three approaches to handling descriptor variables: as single variables, as summed indices, and as profiles. Comparison of these three approaches indicated two advantages of profile models. They facilitated simultaneous consideration of several descriptor variables, and they did not result in loss of information due to averaging (as was the case with summed indices). The discussion of sample aggregation identified a second characteristic of models - a model may emphasize description of the total sample, individuals, or groups.

The final subsection discussed one approach to modeling groups, numerical taxonomy. By analysing variables dealing with types, sources, and time, numerical taxonomic procedures could be used to identify groups representing various prepurchase search processes.

2.3 PREPURCHASE INFORMATION GATHERING AND MARKET SEGMENTATION

The general purpose of this dissertation was development of a descriptive model of prepurchase information gathering. This involved identification of major information gathering styles, and identification of major dimensions differentiating these styles. One rationale for this study was that search styles could be used as a basis for market segmentation.
A paradigm of market segmentation approaches was developed to put into focus search style segmentation. The first characteristic considered was the nature of buyer interests used for segmentation - product-interests or process-interests. The second characteristic was the timing of segment identification - prior or posterior to data analysis. The buyer interest categorization recognized the possibility of segmenting potential buyers based either on their interest in particular product features, or on their interest in using a particular prepurchase process. A simple example of segmentation based on product-related interests would be the design of two product lines, one for buyers interested in economy, and another for those interested in luxury. An example of process-related segmentation would be developing one marketing strategy for buyers interested in shopping at discount stores, and another for those interested in making their purchase at specialty shops. Segmentation could include both product and process interests. The purpose of this categorization was to emphasize the two types.

The purpose of the categorization based on the timing of segment identification was to recognize two possible research approaches. Prior segmentation would assume that the segments of interest could be established on the basis of theory, intuition, or prior research.
The objective of market research would be to gather statistics to describe each segment in terms of factors such as demographics. Posterior segmentation would assume that useful market segments could not be recognized, and thus the objective of market research would be both to identify and describe appropriate segments.

This paradigm resulted in four classes of segmentation. Prior/product-related would include segmenting the beer market into lager and ale drinkers. An example of prior/process-related segmentation would be the "shopping, specialty, convenience" segments suggested by Copeland (1923). Posterior/product-related segmentation could be accomplished by gathering similarities data on the products in a particular market, and then identifying the product interests and market segments via nonmetric multidimensional scaling. An example of the final segmentation approach, posterior/process-related, would be the model discussed earlier, numerical taxonomic identification of search styles.

After categorizing search style segmentation as a process-related approach, it was useful to review earlier writings considering this approach. The first writing appeared in Volume One of the Harvard Business Review, 1922-23. Melvin T. Copeland (1923) suggested that products could be classified on the basis of consumer buying habits. He felt that products fell into
three categories: "convenience", goods purchased at the most convenient store; "shopping", goods for which the consumer desired to compare prices, quality, etc.; and "specialty", goods with some special attraction that induced the consumer to put forth special effort to visit a particular store.

These three categories were elaborated by L.P. Bucklin (1963). He suggested that different buyers may consider a particular product to be in different categories. For example, cameras may be a shopping good for one group of buyers and a specialty good for another. His second suggestion was that retail outlets may be categorized using the same three types. That is, one type of outlet may be based on convenient location, another may carry a broad product line to facilitate shopping within a single store, and another may offer only specialty brands. Since a particular product may be treated as any one of the three types, and a store may use any one of the three approaches to selling the product, Bucklin suggested a three by three classification as a useful approach to market segmentation.

A major similarity between search style segmentation and the segmentation suggested by Copeland was that both assumed discernible patterns of prepurchase activity. Search style segmentation also assumed Bucklin's idea of more than one pattern for a given product. While both
Copeland and Bucklin were primarily interested in pre-purchase patterns with respect to type of outlet, search style segmentation also considered other activities (in particular types of information gathered, sources of information, and time spent). Finally, there was a major distinction in approach between the segmentation suggested by Copeland and Bucklin, and search style segmentation using numerical taxonomy. The earlier work assumed prior segments; the numerical taxonomic approach was based on posterior segment identification.

This section identified how modeling pre-purchase information gathering related to other approaches of market segmentation. In particular, numerical taxonomic identification of search styles was categorized as segmentation based on process-related interests. It was also categorized as an approach where segment identification was a major research objective. Examples of the implication of search style segmentation for marketing management are discussed in the final chapter.

2.4 PREPURCHASE INFORMATION GATHERING: SUMMARY

The discussion of pre-purchase information gathering centered on: what the area was, how a descriptive model of the area could be developed, and why the area was of interest. Three concepts were used to describe the area. First, information gathering was described as multi-dimensional. Factors considered necessary to the
description of the search process were: types of information considered, sources of information considered, and time spent. The second concept discussed was the limited nature of the process. That is, buyers consider a limited number of types and sources, and spend a limited period of time. The third concept, search style, was defined as a particular combination of types, sources, and time. It was introduced to recognize differences across buyers in the extent to which various search dimensions were limited. Consideration of these three concepts suggested that a model of prepurchase information gathering would concentrate on two aspects: identification of major information gathering styles, and identification of dimensions on which styles differed.

The discussion of modeling prepurchase information gathering first considered differences in modeling approaches with respect to level of measurement aggregation and level of sample aggregation. The three measurement categories suggested were models using single descriptor variables, those using summed indices, and models using descriptive variable profiles. The three sample categories suggested were: description of individuals, the total sample, and subsample groups. The final consideration with respect to modeling was a brief introduction to numerical taxonomy. This approach
to modeling was identified as a profile/group model, and was considered to be well matched with the interest in modeling information gathering styles.

The final section discussed prepurchase information gathering in terms of market segmentation. In this discussion numerical taxonomic identification of search styles was described as segmentation based on process-related interests (as opposed to product-related). It was also described as segmentation where the objective was segment identification and description (as opposed to simple segment description).
CHAPTER 3
NUMERICAL TAXONOMIC PROCEDURES

The purpose of numerical taxonomy is to identify subgroups within a general group of subjects, so that the "likeness" of subjects within each subgroup is high, and the "likeness" of subjects between subgroups is low. The term taxonomy refers to a set of interrelated classes or taxons. A simple example demonstrating the interrelated nature is presented below.

Automobiles

Foreign Made               Domestic Made

4- Cylinder                6- Cylinder
6- Cylinder               4- Cylinder
6- Cylinder

The term typology is also used to describe classification schemes. It is used when the interrelationships between classes are not important.

Numerical taxonomic procedures were first developed in the natural sciences, and were used to aid in the classification of plants and insects (Sokal, 1966). Prior to their development taxonomists were forced to base their classification on a limited number of characteristics. Numerical taxonomy made possible classification based on many characteristics.
The application of numerical taxonomic procedures involves three general steps. First, each subject is compared with every other to assess intersubject "likeness" or "proximity". Second, the proximity measures are analyzed using one of several possible cluster analysis algorithms. The final step is to identify the meaningful clusters. This step is of major importance since the descriptive and inferential statistics associated with taxonomic procedures are not well developed. A more complete discussion of these three steps is presented in the following sections.¹

Two terms require attention. Numerical taxonomy and cluster analysis have been used interchangeably in the marketing literature. To avoid confusion, a distinction is made between them in this dissertation. Numerical taxonomy is used as a general term that includes the three steps mentioned above, proximity measures development, cluster identification, and analysis of meaningfulness. Cluster analysis or clustering is used when referring to the second step.

cluster identification.

3.1 PROXIMITY MEASURES

The first step in numerical taxonomic analysis is the comparison of every pair of subjects to determine their proximity. The term proximity is used to include several intersubject comparison measures. Similarity is one type of proximity measure. An example of this type is: comparison of subjects across a set of characteristics to determine the proportion of matching characteristics for each pair of subjects.

\[ S_{ij} = \frac{\text{number of matching characteristics}}{\text{total number of characteristics}} \]

where \( S_{ij} \) is the similarity between subjects \( i \) and \( j \).

The greater the number of matching characteristics, the greater the similarity. Checking for matching scores does not require interval or ordinal assumptions, and thus, this type of proximity measure is particularly suited for use with nominal data. Joyce and Charron (1966) discuss several other similarity measures.

The second type of proximity measure is a distance measure. For example, when the variables are interval measures, the distance between subjects can be calculated using euclidean distance.
\[ D_{ij} = \left[ \sum_{r=1}^{m} (X_{ri} - X_{rj})^2 \right]^{\frac{1}{2}} \]

where \( D_{ij} \) is the euclidean distance between subject \( i \) and \( j \); \( X_{ri} \) is the score of subject \( i \) on dimension \( r \); and distance is being measured across \( m \) dimensions.

In this case the greater the "likeness" the smaller the distance. Green and Rao(1969), and Morrison(1967) discuss other distance measures.

The third type of proximity measure is correlation. Pearson product-moment, Spearman rank, and phi coefficients are examples. The coefficient selected depends on whether the variables being used are interval, ordinal or nominal. The product-moment equation is present as an example.

\[ r_{ij} = \left[ \sum_{r=1}^{m} \frac{(X_{ri} - \bar{X}_i) (X_{rj} - \bar{X}_j)}{S_i S_j (m-1)} \right]^{\frac{1}{2}} \]

where \( r_{ij} \) is the product moment correlation between subject \( i \) and \( j \); \( X_{ri} \) is the score of subject \( i \) on dimension \( r \); \( \bar{X}_i \) and \( S_i \) are the mean and standard deviation of the scores of subject \( i \); and \( m \) is the number of variables being considered.

An important characteristic of correlation is the removal of the mean and standard deviation (level and scatter) of each subject's scores. For example, comparing subjects based on several personality traits, two subjects can correlate highly even when one subject
has all low trait scores, and the other has all high ones. Similarly, high correlation between subjects can occur when the trait scores of one subject have a large standard deviation, and those of another subject have a small one.

The type of proximity measure used is a function of two factors. The first is the nature of measurement. Similarity or phi correlation can be used with nominal data, Spearman's rank correlation with ordinal data, and euclidean distance or product moment correlation with interval data. When using interval measures a second factor to be considered is the importance of differences across subjects in the level and scatter. When level and scatter are not important correlation is used. When they are important a distance measure is used.

The final consideration in the development of proximity measures is to insure each variable has the possibility of equal impact. In two situations this may not be the case, non-equivalent variable scales, and intercorrelated variables. The effect of non-equivalent scales can be seen by considering the potential differences in impact of two scales, one with a range of 0 to 1 and the other with a range of 10 to 20. Based on these two scales the distance between subjects will be determined largely by the 10 to 20 variable. Con-
sideration of the product-moment equation indicates that non-equivalent scales are also a problem when using correlation. To overcome this problem each variable can be standardized to zero mean and unit variance prior to calculation of the proximity measure (Frank and Green, 1968).

Intercorrelated variables are the second cause of unequal impact. For example, consider the situation where the proximity measure is based on three variables; two variables have a correlation of 1.0, and the other correlations are 0.0. In effect only two dimensions are being measured; however, since one dimension is being measured twice it will have twice the impact on the proximity measure. To overcome this problem principle components analysis can be used to develop orthogonal dimensions prior to proximity calculation (Frank and Green, 1968).

3.2 CLUSTERING

The second step in numerical taxonomic analysis is segmentation of the proximity matrix into clusters. There are several approaches to clustering. In this discussion they are categorized into three types; however, the distinction is not always as clear as is suggested here. One approach is inverse factor analysis, or Q-factoring. This usually begins with a correlation
matrix. The matrix is factored into hypothetical types, and the subject-clusters are based on subjects' factor loadings on the hypothetical types (Johnson, R.M., 1970).

A second clustering approach, the space-density approach, positions the subjects in a space defined by a number of descriptor variables. This space is then segmented into hypercubes and the number of subjects appearing in each hypercube are counted. Clusters are identified based on concentrations in the subject-scatter (Myers and Nicosia, 1968).

The final approach, is hierarchical (or linkage) clustering. This approach identifies clusters based on a particular criterion. For example, a clustering criterion could be: subjects joining a cluster must have a proximity of a given level or better with every subject already in the cluster. Another example of a clustering criterion is: subjects joining the cluster must have a similarity of a given level or better with any other subject in the cluster. The former criterion is called complete linkage or diameter\(^2\) method; the latter is called single linkage.

Using the diameter method, clusters are based on likeness across all subjects in a cluster. Thus, this

\(^2\) Cluster diameter is defined as the maximum distance (or minimum similarity) between any pair of subjects in the cluster.
approach leads to the identification of relatively compact (spherical) clusters. Using single linkage, clusters are based on likeness between any pair of cluster members. This approach can result in a "chain-type" cluster. For example, "A" may cluster with "B" based on one subset of characteristics, and "B" with "C" on another subset. Thus, although "A" and "C" are in the same cluster, they have no common characteristics.

The hierarchical clustering process sequentially reduces the criterion level required for cluster membership. Thus, at the start of the clustering scheme each subject is in a "cluster of one"; at the end all subjects are in one cluster.

As an example of the clustering process, Johnson's "diameter" algorithm (Johnson, S.C., 1967) is applied to a simple example, Figure 3.1. The stimuli for the example are plotted in Figure 3.1(a). Clustering of the stimuli begins with the euclidean distance matrix shown in Figure 3.1(b). The first step is to identify the minimum distance appearing in the matrix (1.00). This is the first criterion level at which clustering can occur. At this level clusters are formed by linking (or joining) 1 to 4, and 5 to 6. The next step is to transform the distance matrix by replacing the linked stimuli by the new cluster, Figure 3.1(c). The distance entry between a new cluster and another stimulus is the
FIGURE 3.1
EXAMPLE OF CLUSTERING PROCESS USING
JOHNSON'S "DIAMETER" ALGORITHM

(a) Stimuli Location

(b) Euclidean Distance Matrix
(* minimum distance)

(c) Distance Matrix Transformation One

(d) Distance Matrix Transformation Two

Criterion Levels

Criterion Level Stimuli Number

1 4 2 3 5 6

0.00 . . . . .
1.00  x  x  x  .  x  x  x
2.23  x  x  x  x  x  x  x  x  x  x
5.38  x  x  x  x  x  x  x  x  x  x

(e) Clusters at Three Criteria Levels

(f) Clustering: "Diameter" Algorithm Output
maximum distance between that stimulus and the cluster members prior to clustering. That is, the distance 1-4 to 2 is the maximum of 1 to 2 and 4 to 2 (2.33). This method of distance matrix transformation is central to the diameter clustering algorithm. It insures that a stimulus will not be linked to a cluster until the criterion level has been reduced to the maximum distance between the stimulus and any cluster member.

The process is continued by searching for the next minimum distance in the transformed matrix (2.33). This is the second criterion level. At this level 2 is linked to 1-4 and 3 is linked to 5-6. Once again the distance matrix is transformed, Figure 3.1(d). At this point there is only one intercluster distance remaining, so that at the next criterion level all stimuli will be in one cluster. Figure 3.1(e) represents the stimuli clustered at the three criterion levels, and Figure 3.1(f) is an example of the output provided by the Johnson algorithm.

There are two advantages of hierarchical clustering generally not available with the other two approaches. First, only rank order assumptions are required of the proximity matrix. Second, consideration of cluster characteristics at various levels in the hierarchical scheme provides useful information for identification and definition of relevant clusters.
3.3 ANALYSIS OF CLUSTERING RESULTS

The final step in the application of numerical taxonomy is the analysis of clustering results. This involves deciding on the final set of clusters, and assessing the "meaningfulness" of the results. Because of the relative infancy of numerical taxonomy, there are no well developed procedures for this phase of the analysis. However, there are several criteria that provide guidelines. On this point Shuchman (1966) criticizes Green, Frank and Robinson's (1967) identification of similar test markets. He quarrels with their conclusion that numerical taxonomic procedures are helpful in a variety of marketing problems. He feels that their work provides no evidence of usefulness, or internal validity. Although Shuchman's stand appears to be extreme, with the current state of the art considerable judgement is required in the evaluation of clustering results.

The general purpose of numerical taxonomy is to establish clusters so that members of a particular cluster are more similar to others in the same cluster than they are to non-members. On these grounds Joyce and Charron (1966) suggest a useful criterion for deciding the appropriate number of clusters is comparison of average similarity of subjects within and between clusters. When the number of clusters is large the
average within cluster similarity will be high, and the between-cluster similarity will be relatively low. As the number of clusters is reduced the difference between these two averages will decrease, until at some point the researcher no longer feels there is any meaningful difference. "Between-within" comparison provides a guide for stopping the hierarchical clustering; however, judgement remains a critical ingredient.

Assessment of cluster "meaningfulness" involves describing the clusters, assessing cluster reliability, and assessing cluster validity. Cluster centroids are the usual method of describing clusters, cluster centroid being the cluster average on each variable used in the proximity matrix calculation. For example, if proximity is based on three variables, A, B, and C, a cluster centroid is the profile of A, B, and C (averaged across cluster members). Two additional statistics useful in describing clusters are average within-cluster proximity and cluster diameter. These provide a guide to the relative cluster size.

Two techniques are available for assessing the reliability of the clusters. First, use of more than one taxonomic approach provides an indication of the degree to which cluster results are dependent on numerical taxonomic procedures. Joyce and Charron (1966), and Frank and Green (1968) emphasize comparison of
alternative approaches; Green and Rao (1969) discuss a measure for assessing the similarity of clusterings resulting from different approaches. Second, split-half analysis can be used to assess cluster reliability. For this type of analysis the sample is divided randomly into two groups, and parallel analyses are performed. This provides an indication of the extent to which cluster-types are independent of the sample under consideration.

A final approach for assessing cluster meaningfulness is by attempting to establish concurrent validity by relating cluster membership to external variables. Kernan (1968) used this approach, and showed that the cluster profiles on the characteristics used in proximity development judgamentally matched profiles based on external variables. An example of a more formal approach to concurrent validation is to explain cluster membership in terms of external variables using multiple discriminant analysis.

3.4 NUMERICAL TAXONOMIC PROCEDURES: SUMMARY

The purpose of numerical taxonomic analysis is to segment a sample into clusters so that the "likeness" of subjects within each cluster is high, and between clusters is low. The first step in the application of these procedures is the development of a proximity
matrix indicating the degree of "likeness" of every pair of subjects. Three types of proximity measures are available: similarity, distance and correlation. The measure selected is based on the nature of measurement, and the importance of differences between subjects in level and scatter of their scores. The final consideration with respect to proximity development is the potential impact of each variable on the proximity measure. Two conditions lead to unequal impact - variables with scale differences and intercorrelated variables. Scale differences can be rectified by standardizing each variable prior to proximity calculations. Correction for intercorrelation can be accomplished by using principle components analysis to develop orthogonal dimensions prior to proximity calculations.

The second step in numerical taxonomic analysis is segmenting the proximity matrix into clusters. Three approaches were described, inverse factor analysis, space-density clustering, and hierarchical clustering. The latter provides two advantages generally not present in the other two. It requires only rank-order assumptions of the proximity matrix. Second, it identifies the linkage between small and large clusters, information of value when assessing cluster meaning.
The final step involves determining the appropriate number of clusters, describing the clusters, and assessing their reliability and validity. Because of weaknesses in the descriptive and inferential statistics associated with numerical taxonomic procedures, this area ("meaningfulness") requires considerable attention. Several procedures provide a guide. Comparison of the average proximity between and within clusters is useful in determining the appropriate number of clusters. Cluster centroids, average cluster proximity and cluster diameter are used as descriptive statistics. Reliability is assessed by comparing the clustering results obtained using alternative taxonomic procedures, and by using split-sample analyses. Finally, relating clustering results with external variables is a useful approach to concurrent validation.
CHAPTER 4
RESEARCH APPROACH AND DATA BASE

4.1 AREAS OF STUDY

The central proposition of the research was:

The analysis of measures of information gathering by means of numerical taxonomic procedures will result in progress toward a general model of pre-purchase information gathering. In particular, this approach will lead to identification of major pre-purchase information gathering styles, and identification of major dimensions differentiating these styles.

This proposition identifies the three areas of study: measures of information gathering, numerical taxonomic procedures, and modeling prepurchase information gathering. Each area is discussed below.

4.1.1 Measurement of Prepurchase Information Gathering

The definition of prepurchase information gathering used in this research was: the set of activities directed toward gathering information for purposes of a purchase decision. Measurement of information gathering involved three areas: types of information gathered, sources of information considered, and time spent. One question was central, how could these areas be measured to facilitate the identification of groups of buyers with similar information gathering styles?
4.12 Numerical Taxonomic Procedures

The purpose of numerical taxonomy was to segment a sample into subject clusters, so that the "likeness" of subjects within each cluster was high, and the "likeness" of subjects between clusters was low. As an area of study two factors were considered. First, what proximity measures and clustering algorithms were consistent with the variables being used and the objectives of the study. Second, how could clustering results be evaluated.

4.13 General Model of Prepurchase Information Gathering

The long-run purpose underlying the research was the development of a general model of prepurchase information gathering. General model was intended to mean a model that describes the information gathering of a broad range of buyers, purchasing a broad range of products. The objective of this research was to progress toward a general model by attempting to describe household durable buyers in terms of major information gathering styles and major dimensions differentiating these styles. The term major provided an indication of the desired level of generality. The generality of the dimensions was a level similar to Katona and Mueller's "five dimensions of deliberation". They judgementally selected "extent of circumspectness",
"extent of information seeking", "choice with respect to price", "choice with respect to brand", and "number of features considered" as major descriptive dimensions. For this research the approach was to extract major dimensions from the data, rather than to form them based on judgement. The level of generality, however, was similar.

The term style was defined in Chapter 2 as a particular information gathering process, or a particular combination of types, sources and time. Thus, major styles, referred to a particular information gathering process used by a reasonable proportion of the sample. The Katona and Mueller study provided a guide to "a reasonable proportion". Mueller(1955, page 53) identified three buyer groups that varied in degree of deliberation (high, moderate and low), and several "feature substitution" groups. This research concentrated on styles representing a similar proportion of the sample.

The questions being addressed in this area of study were first, to what extent could major information gathering styles be identified? Second, what dimensions differentiated these styles? Third, if continuing work on information gathering styles and dimensions seemed useful, how should future research proceed?
4.2 GENERAL RESEARCH APPROACH

The research was considered exploratory for two reasons. First, identifying buyer groups based on information gathering styles was relatively untested. Second, numerical taxonomy required special attention with respect to interpretation of clustering results. Thus, rather than test specific hypotheses, the objective was to make progress in the three areas included in the central proposition - measurement, numerical taxonomy, and modeling information gathering.

Since the research was not based on hypothesis testing, care to avoid reporting chance findings was particularly important. Two major precautions were taken in this regard. First, at every step in the research, parallel analyses were done on two random halves of the sample. Comparison across halves provided an indication of the reliability of the findings. The second precaution was the sequential analysis of two product samples, furniture buyers and appliance buyers. Although some differences due to product were expected, similarity across samples reduced the possibility of chance findings.

4.3 METHODOLOGY

4.3.1 Data Base

The data used in this research was part of a consumer survey designed by Professors J.N. Fry and B. Portis of the School of Business Administration,
University of Western Ontario. The focus of the survey was prepurchase information gathering of buyers of household durables, such as living room suites, refrigerators, and televisions. Earlier research done by George Katona and Eva Mueller of the Survey Research Center, attempted to...

...gain an understanding (a) of "deliberation" in purchasing and (b) of the nature of alternative processes of decision making (Katona and Mueller, 1955).

The Fry-Portis study was designed to replicate and extend the Katona-Mueller study.

The data was collected by means of a questionnaire administered via personal interview. The interviews were conducted by a commercial research organization, Canadian Facts Ltd. A stratified quota sample was gathered consisting of 546 London, Ontario households. Stratification was used to insure representation from major socio-economic areas in the city. Each respondent was a female head of household, and had purchased within 24 months prior to the survey one of 13 large household durables. Four furniture items were considered: living room suite, upholstered chair, dining room suite, and bedroom suite. Nine appliances were considered: washing machine, clothes dryer, stove, refrigerator, freezer, color television, air conditioner, black and white television, and dish washer. If the respondent had
purchased more than one of these items within the 24 month period, she was questioned only with respect to the most recent purchase.

The survey gathered four types of data. The first type dealt with "how" information was gathered, for example information sources used and time spent on the decision. The second type dealt with "why" information gathering was handled the way it was. The third type was data on enabling and constraining factors such as perceived product complexity and condition of old product. The final type was socio-economic data.

This research was based on the data describing how information was gathered. Five aspects of the information gathering process were measured. Four questions dealt with the type and range of alternatives considered. Nine dealt with the information sources used. Two dealt with the features considered. One dealt with the stores visited, and one question determined the time spent considering the purchase. In terms of the conceptual model of information gathering suggested in Chapter 2, alternatives and features measured type of information considered; sources and visits measured sources considered; and time was measured directly. Figure 4.1 lists the questions used to obtain these data.
FIGURE 4.1
QUESTIONS MEASURING INFORMATION GATHERING

Alternatives

7.2 Before you decided to buy the ___________, did you consider saving the money?

7.3 Before you decided to buy the ___________, did you consider spending the money on something else?

10. Did you know from the beginning what brand or make of furniture you wanted, or did you consider a few different brands, or was it a wide-open choice?

14.4 Before deciding on the ____________ you bought did you also consider other which cost much more or much less than what you bought?

Information Sources

13.1 Did you get any useful information from friends or relatives?

13.3 Did you get any useful information from salesmen?

13.5 Did you get any useful information from trips to stores?

13.7 Did you get any useful information from advertisements?

13.9 Can you recall any other source of information which was helpful in your shopping? (PROBE FOR BOTH KIND OF INFORMATION AND SOURCE)

15.4 Did you phone any stores or private sources not listed above regarding the ____________?

16. Before buying, did you see a ____________ like the one you bought at a friend's or relative's house?

17. Did you discuss your planned purchase of a ____________ with any relative, friend, to get their experience or ideas?

17.1 If Yes, would you say you discussed the planned purchase with friends or relatives at length or only a little?

18. Did you discuss your planned purchase of a ____________ with your family at length or only a little?

Features

11.5 What other qualities made a difference?

12. What were the main features, for example, size, style, price, strength, store you considered in purchasing the ____________?

Store Visits

15. We are interested in finding out about the place where the purchase was made. Please consider each store or person selling goods privately as a different place of purchase.

What stores or other places of purchase did you visit while shopping for the ____________?

Time

7. Concerning your recent purchase of the ____________ how long were you thinking or talking about buying it before you actually made the purchase?
4.32 Sequence of Analysis

The analysis in this research was composed of three general segments. The first was evaluation of the Fry-Portis measures of prepurchase information gathering. The second was assessment and development of numerical taxonomic procedures. The final segment was numerical taxonomic analysis of the information gathering data. A flow chart of the research is presented in Figure 4.2.

A preliminary evaluation of the Fry-Portis measures of prepurchase information gathering was done to establish a set of variables for numerical taxonomic analysis. Each measure was evaluated for potential response bias, discrimination across subjects, and potential redundancy. This analysis is reported in Chapter 5.

The preliminary analysis of numerical taxonomic procedures selected proximity measures and a clustering algorithm that were compatible with the data. This segment of the analysis also evaluated existing means of analyzing clustering results, and developed two new procedures useful for this purpose. This is reported in Chapter 6.

The final segment, numerical taxonomic analysis of information gathering data, consisted of exploration and re-test analyses. The exploration analysis was based on data describing 287 furniture buyers. The approach was to test a variety of combinations of
FIGURE 4.2
RESEARCH DESIGN

Fry-Portis Data
17 Measures of Information Gathering
Furniture Sample: 287 Subjects
Appliance Sample: 259 Subjects

Preliminary Analysis
N/T Procedures
- select proximity measures
- select clustering algorithm
- assess and develop cluster evaluation procedures
Final Set of N/T Procedures

Preliminary Analysis
Information Gathering Measurement
- response biases
- across-subjects discrimination
- measurement redundancy
- nominal and interval variables
Final Set of Variables

Exploration Analysis
Split Sample Comparisons
Group A | Group B
Numerical Taxonomic Analysis
Iterative Trials of Several Variable - Procedure Combinations

"Best" Variable - Procedure Combinations

Dimensions and Styles
Furniture Buyers

Dimensions and Styles
Appliance Buyers

Re-test Analysis
Split Sample Comparisons
Group A | Group B
Numerical Taxonomic Analysis
variables and procedures. The objectives were first, to search for major styles used by the furniture buyer sample, and second, to identify the "best" combination(s) of variables and procedures. "Best" was indicated by relative success in style identification. The re-test analysis was based on data describing 259 appliance buyers. The objective of this analysis was to use the "best" combinations to identify major styles used by the appliance buyer sample.
CHAPTER 5
MEASUREMENT OF PREPURCHASE INFORMATION GATHERING
PRELIMINARY ANALYSIS

Questions describing prepurchase information gathering were selected from the Fry-Portis survey. They were categorized in five groups: alternatives considered, information sources used, features considered, stores visited, and time spent. Each group was evaluated to establish a set of variables for numerical taxonomic analysis. This preliminary analysis was done using data describing furniture buyers. For split-sample comparisons, prior to preliminary analysis the furniture sample was randomly divided into halves, "Group A" and "Group B".

There were three criteria central to the evaluation of each variable. First, consideration was given to the wording of the question to assess possible response biases. For example, would the question wording suggest the "right" answer? Second, the distribution of responses was studied to determine the degree to which the variable discriminated across subjects. A variable that had similar responses for all subjects was of no
value in distinguishing one subject from another. Third, correlation between variables was considered to assess areas of redundant measurement.

A variable assessed as very weak on any of these criteria was eliminated. The variables remaining were termed the final variable set, and used for numerical taxonomic analysis.

In addition to evaluating individual variables a major concern in the preliminary analysis was to consider alternative forms of variables, non-aggregated and aggregated. As indicated in Figure 4.1 most of the questions had dichotomous response categories. It will be seen subsequently that using each question as an individual variable the final set included 18 nominal variables. An alternate approach was to develop a smaller number of interval variables, "number of alternatives", "number of sources", "number of features", etc. Clearly, information was lost using the interval approach. For example, the nominal approach to "information sources" resulted in a set of variables containing information as to how many sources and which ones. On the other hand the interval approach resulted in a single variable which only contained information as to how many sources.

Even though this loss of information occurred, the interval approach remained of interest for two reasons.
First, it was not clear whether non-aggregated or aggregated data was more appropriate for numerical taxonomic analysis of information gathering. Second, the interval approach simplified interpretation of information gathering styles. A style consisted of a profile on a series of variables, and thus, the complexity of interpretation was related to the number of variables being considered. The preliminary data aggregation associated with the interval approach, therefore, simplified prepurchase style interpretation.

5.1 EVALUATION OF VARIABLES: ALTERNATIVES

There were four questions in the Fry-Portis survey questionnaire measuring alternatives: (1) question 7.2 addressed the consideration of saving rather than purchase; (2) question 7.3, spending on an alternate product; (3) question 10, alternative brands; and (4) question 14.4, alternative prices. The wording and response distributions for these questions is presented in Figure 5.1. The correlations between the four variables for the split-half subsamples is presented in Table 5.1. Product moment correlation was calculated throughout the preliminary assessment of variables. For the dichotomous variables (all but five) this was equivalent to calculating the phi coefficient. For the two trichotomous variables, alternative brands and
discussion with friends, neither product moment nor phi coefficient were correct for inferential purposes. However, as a guide to measurement redundancy, the use of product moment was judged to be satisfactory. Subjects not responding to one of the variables were removed before correlation calculations. There were only nine non-responses to the alternatives variables.

Consideration of Figure 5.1 and Table 5.1 led to changes in three variables. First, alternate spending displayed very skewed distribution with only 6.6 per cent positive responses. Although the initial reaction to this skewness was to eliminate the variable, conceptual similarity between alternate spending and the saving variable led to combining these two variables to form alternate use of money. The correlation between the spending and saving variables also supported this combination. A subject's response to the new variable was positive if she had a positive response to either or both of the spending and saving variables.

Alternative brands was also modified. When forming a proximity measure using nominal variables it was desirable to have an equal number of categories in each variable. As suggested in the discussion of numerical taxonomic procedures, failure to follow this procedure resulted in unequal weighting of the variables. Since most of the variables were dichotomous, it was desirable
### FIGURE 5.1
MEASUREMENT OF ALTERNATIVES

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Question</th>
<th>Response Distribution</th>
<th>Group A</th>
<th>Group B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.2</td>
<td>Before you decided to buy the blank did you consider saving the money?</td>
<td>Yes</td>
<td>44</td>
<td>41</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td>100</td>
<td>102</td>
<td>202</td>
</tr>
<tr>
<td>7.3</td>
<td>Before you decided to buy the blank did you consider spending the money on something else?</td>
<td>Yes</td>
<td>7</td>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td>137</td>
<td>131</td>
<td>268</td>
</tr>
<tr>
<td>10.</td>
<td>Did you know from the beginning what brand or make of furniture you wanted, or did you consider a few different brands, or was it a wide-open choice?</td>
<td>Only one brand considered</td>
<td>11</td>
<td>10</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A few brands considered</td>
<td>20</td>
<td>15</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wide-open Choice</td>
<td>113</td>
<td>116</td>
<td>229</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No response</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>14.4</td>
<td>Before deciding on the blank you bought did you also consider other blank which cost much more or much less than what you bought?</td>
<td>Yes</td>
<td>56</td>
<td>59</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td>82</td>
<td>83</td>
<td>165</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No response</td>
<td>6</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Variable</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>1 Saving</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.00b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Spending</td>
<td>0.20</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Brand</td>
<td>-0.10</td>
<td>-0.00</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>-0.06</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Price</td>
<td>-0.10</td>
<td>-0.02</td>
<td>-0.08</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.17</td>
<td>0.14</td>
<td>-0.17</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

* The correlations are phi-coefficients except when "brand" is one of the variables being correlated, in which case product moment correlation is calculated.

* The top correlation is for Group A; the bottom, Group B.
to change alternative brands to a dichotomous format. The response pattern for this variable indicated that 80 percent of furniture purchasers considered brand a "wide open choice". Thus, alternate brand was modified forming a dichotomous variable, wide-open choice or one-few brands.

In summary the final set of variables measuring alternatives were: alternative use of money, alternative brands, and alternative prices. All three variables were dichotomous, and were used directly for the nominal approach. The interval approach was accomplished by forming one variable, "number of alternatives". This variable was a count of the number of endorsements to the three nominal variables.

5.2 EVALUATION OF VARIABLES: INFORMATION SOURCES

There were nine questions addressing information sources used. The first considered information from friends and relatives; second, from salesmen; third, from stores; fourth, from advertisements; and fifth, from "other" sources. The sixth question considered "seeing one like the one you bought at friends"; seventh, discussion with friends; and eighth, discussion with family. The final question addressed the use of telephone as a potential source of information. The wording of these questions and the response pattern for each
appears in Figure 5.2. There were 20 non-responses to the source questions; they were removed before the correlation calculations, Table 5.2.

Three changes were made in the information source variables. First, use of telephone was dropped because of its skewed response distribution (only 7 percent positive endorsement), and because conceptually it was slightly different from the other source variables. Use of telephone described how the information was obtained, rather than identifying the source of the information. The two other changes were dropping of discuss with friends and relatives, and dropping of see at friends and relatives. These variables displayed consistent intercorrelations, and were both consistently correlated with information from friends. In order to eliminate unnecessary redundancy only information from friends was retained. This variable was selected because it had similar wording to several other source variables.

In summary the final set of information source variables comprised the following: information from friends, salesmen, stores, advertisements, other, and discussion with family. Each of these variables was dichotomous, and was used directly for the nominal approach. For the interval approach positive endorsements across these six variables were counted to form "number of information sources used".
<table>
<thead>
<tr>
<th>Question Number</th>
<th>Question</th>
<th>Response Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.1</td>
<td>Did you get any useful information from friends or relatives?</td>
<td>Yes: 27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group A</td>
</tr>
<tr>
<td>13.3</td>
<td>Did you get any useful information from salesmen?</td>
<td>Yes: 52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group A</td>
</tr>
<tr>
<td>13.5</td>
<td>Did you get any useful information from trips to stores?</td>
<td>Yes: 87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group A</td>
</tr>
<tr>
<td>13.7</td>
<td>Did you get any useful information from advertisements?</td>
<td>Yes: 52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group A</td>
</tr>
<tr>
<td>13.9</td>
<td>Can you recall any other source of information which was helpful in your shopping? (PROBE FOR BOTH KIND OF INFORMATION AND SOURCE)</td>
<td>Yes: 19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group A</td>
</tr>
<tr>
<td>15.4</td>
<td>Did you phone any stores or private sources not listed above regarding the _____?</td>
<td>Yes: 13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group A</td>
</tr>
<tr>
<td>16.</td>
<td>Before buying, did you see a like the one you bought at a friend's or relative's house?</td>
<td>Yes: 17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group A</td>
</tr>
<tr>
<td>17.6</td>
<td>Did you discuss your planned purchase of a _____ with any relative, friend, to get their experience or ideas?</td>
<td>No: 110</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group A</td>
</tr>
<tr>
<td>17.1</td>
<td></td>
<td>At length: 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>At length: 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No response: 11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total: 105</td>
</tr>
<tr>
<td>18.</td>
<td>Did you discuss your planned purchase of a _____ with your family at length or only a little?</td>
<td>At length: 51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total: 105</td>
</tr>
</tbody>
</table>
TABLE 5.2
MEASUREMENT OF INFORMATION SOURCES
PHI CORRELATION MATRIX\textsuperscript{a}

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Info. Friends</td>
<td>1.00\textsuperscript{b}</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Info. Salesmen</td>
<td>0.00</td>
<td>1.00</td>
<td>-0.12</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Info. Stores</td>
<td>-0.19</td>
<td>0.12</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Info. Ads</td>
<td>0.12</td>
<td>0.03</td>
<td>0.05</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Info. Other</td>
<td>0.02</td>
<td>-0.04</td>
<td>0.19</td>
<td>0.09</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phone</td>
<td>0.12</td>
<td>0.11</td>
<td>0.08</td>
<td>0.20</td>
<td>0.01</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>See at Associates</td>
<td>0.49</td>
<td>-0.06</td>
<td>-0.14</td>
<td>0.04</td>
<td>0.08</td>
<td>0.08</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discuss Associates</td>
<td>0.61</td>
<td>-0.09</td>
<td>-0.05</td>
<td>0.13</td>
<td>-0.01</td>
<td>-0.03</td>
<td>0.45</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Discuss Family</td>
<td>0.04</td>
<td>0.07</td>
<td>0.06</td>
<td>0.22</td>
<td>0.13</td>
<td>0.15</td>
<td>0.01</td>
<td>0.10</td>
<td>1.00</td>
</tr>
<tr>
<td>Family</td>
<td>0.12</td>
<td>-0.02</td>
<td>0.04</td>
<td>0.15</td>
<td>0.11</td>
<td>0.09</td>
<td>0.22</td>
<td>-0.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Since "discussion with associates" was trichotomous product moment correlation correlations rather than phi-coefficients were used with this variable.

\textsuperscript{b} Top correlation is for Group A; bottom, Group B.
5.3 EVALUATION OF VARIABLES: FEATURES

The features considered were measured via questions 11.5 and 12. Since these two questions appeared together on the questionnaire they were combined. The responses were coded as eight dichotomous variables: price, brand, store, style, quality, size, color, and service. Wording of the questions, and endorsement frequencies are presented in Table 5.3. The phi correlation matrix is shown in Table 5.4.

There were four feature variables requiring attention because of low endorsement, brand, store, color, and service. Service was dropped immediately since none of the subjects considered it important. Store was endorsed by only six percent of the subjects, and therefore, was dropped. Brand, although endorsed by only three percent, was retained since it was expected to be considerably more important for the re-test sample of appliance purchasers. Finally, the lack of correlation between color and style in spite of their conceptual similarity suggested that they were substitute variables. That is, although they were measuring similar features (physical appearance), the open-ended nature of the question resulted in positive response to either one or the other. Thus, style and color were combined - subjects received a positive score if they had endorsed style, color, or both.
TABLE 5.3
MEASUREMENT OF FEATURES

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Question Wording</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.5</td>
<td>What other qualities made a difference?</td>
</tr>
<tr>
<td>12.</td>
<td>What were the main features, for example, size, style, price, sturdiness, store, you considered in purchasing the __________?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Response Category</th>
<th>Endorsements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group A</td>
</tr>
<tr>
<td>1 Price</td>
<td>85</td>
</tr>
<tr>
<td>2 Brand</td>
<td>7</td>
</tr>
<tr>
<td>3 Store</td>
<td>4</td>
</tr>
<tr>
<td>4 Style</td>
<td>98</td>
</tr>
<tr>
<td>5 Quality</td>
<td>66</td>
</tr>
<tr>
<td>6 Size</td>
<td>72</td>
</tr>
<tr>
<td>7 Color</td>
<td>24</td>
</tr>
<tr>
<td>8 Service</td>
<td>0</td>
</tr>
</tbody>
</table>
### TABLE 5.4

**MEASUREMENT OF FEATURES**

**PHI CORRELATION MATRIX**

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Price</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Brand</td>
<td>-0.12</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.02</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Store</td>
<td>0.02</td>
<td>-0.03</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.09</td>
<td>-0.04</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Style</td>
<td>-0.19</td>
<td>-0.01</td>
<td>-0.01</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.11</td>
<td>0.10</td>
<td>-0.12</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Quality</td>
<td>-0.09</td>
<td>-0.13</td>
<td>0.06</td>
<td>0.03</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.01</td>
<td>0.12</td>
<td>-0.06</td>
<td>-0.13</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Size</td>
<td>-0.05</td>
<td>-0.07</td>
<td>-0.05</td>
<td>-0.07</td>
<td>0.02</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td>-0.10</td>
<td>-0.06</td>
<td>0.10</td>
<td>-0.08</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>7 Color</td>
<td>-0.03</td>
<td>0.01</td>
<td>-0.07</td>
<td>-0.02</td>
<td>-0.08</td>
<td>0.02</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>-0.01</td>
<td>-0.06</td>
<td>0.11</td>
<td>0.02</td>
<td>-0.03</td>
<td>-0.06</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*a* "Service" was not included since it was not endorsed by any of the subjects.

b The top correlations are for Group A; the bottom, Group B.
Thus, five variables made up the final "features" set: price, brand, style, quality, and size. Each variable was dichotomous and was used for the nominal approach. The interval approach variable, number of features considered, was formed by counting the number of nominal variables endorsed.

5.4 EVALUATION OF VARIABLES: STORE VISITS

Question 15 obtained information on the number of stores visited, and the number of visits to each store. Wording for question 15 and response distributions for number of stores visited, total visits, and maximum visits to a single store is presented in Table 5.5. Maximum visits was included to differentiate subjects who visited many stores a few times from those who visited a few stores many times. These two subject types intuitively represented different decision processes. For example, subjects with visits-by-store distributions of 5-1-1-1 and 2-2-2-2 have visited the same number of stores (four) and have the same total visits (eight). However, they appeared to characterize different styles, possibly "procrastination" and "thoroughness". The product moment correlation matrix for the three store variables is presented in Table 5.6.

The interval nature of the store visits variables was not compatible with the nominal approach. Considera-
TABLE 5.5
MEASUREMENT OF STORE VISITS

Question Number

15 We are interested in finding out about the place where the purchase was made. Please consider each store or person selling goods privately as a different place of purchase.

What stores or other places of purchase did you visit while shopping for the

How many visits did you make to this place?

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Number of Stores</th>
<th>Number of Visits</th>
<th>Maximum Visits To Single Store</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>45</td>
<td>34</td>
<td>79</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>19</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>24</td>
<td>52</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>24</td>
<td>46</td>
</tr>
<tr>
<td>5</td>
<td>17</td>
<td>17</td>
<td>34</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td>18</td>
<td>31</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
TABLE 5.6
MEASUREMENT OF STORE VISITS
PRODUCT MOMENT CORRELATION MATRIX

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Numbers of Stores</td>
<td>1.00(^a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Total Visits</td>
<td>0.71</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.74</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>3 Maximum Visits</td>
<td>0.35</td>
<td>0.85</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>0.32</td>
<td>0.75</td>
<td>1.00</td>
</tr>
</tbody>
</table>

\(^a\) Top correlations are for Group A; bottom, Group B.
tion of the response distribution for each variable led to the identification of nominal categories. The response distribution for number of stores appeared to be bimodal with the frequency of "2 stores" being relatively low. Therefore, the categories established for this variable were, 2 or less, and 3 or more. The total frequencies for these two categories were 112 and 175 respectively. The second variable, total visits, appeared to display a trimodal distribution with the frequencies of "2 visits", and "5 visits" being relatively low. Three categories were established, 2 or less, 3 to 5, and 6 or more, with total frequencies of 90, 102, and 95. Finally, the response distribution of maximum visits was highly skewed toward few visits. Two categories were formed, 1 or less, and 2 or more, with total frequencies of 166 and 121.

The three nominal variables discussed here represented the initial set for the nominal approach. The high intercorrelations shown in Table 5.5 indicated total visits to be a candidate for early removal during the exploration analysis. On the other hand total visits represented a reasonable summary variable, and thus was selected for the interval approach.
5.5 EVALUATION OF VARIABLES: TIME

Question 7 dealt with time spent on prepurchase decision making. The wording and response distribution for this question is presented in Table 5.7. The relatively uniform response distribution indicated that the variable provided subjects with a reasonable range of "time" categories. Thus, the variable was used without alteration in both the nominal and quantitative approaches.

The weakness of the time variable for use in the interval approach was the necessity of assuming equal intervals between categories. That is, the assumption was that subjects perceived the interval from "a few days" to "several months" as twice the interval from "a few days" to "a few weeks", and that "a few days" to "a few weeks" represented the same interval as "several months" to "several years". Clearly, in the strict sense of the words the categories did not represent equal intervals. However, since the subject was faced with four possible responses, it was probably reasonable to assume that she perceived them simply as four increasing time categories, and implicitly considered them to be of equal interval.

5.6 PRELIMINARY MEASUREMENT ANALYSIS: SUMMARY

Evaluation of the variables in each of five information gathering areas led to the selection of 18
TABLE 5.7
MEASUREMENT OF TIME

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Question Wording</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.</td>
<td>Concerning your recent purchase of the ___ how long were you thinking or talking about buying it before you actually made the purchase?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Response Distribution</th>
<th>Group A</th>
<th>Group B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Several years</td>
<td>43</td>
<td>25</td>
<td>68</td>
</tr>
<tr>
<td>Several months</td>
<td>61</td>
<td>62</td>
<td>123</td>
</tr>
<tr>
<td>A few weeks</td>
<td>23</td>
<td>34</td>
<td>57</td>
</tr>
<tr>
<td>A few days</td>
<td>17</td>
<td>22</td>
<td>39</td>
</tr>
</tbody>
</table>
nominal variables as the final set for numerical taxonomic analysis. Three variables dealt with alternatives, use of money, brands, and price. Six dealt with sources, friends, salesmen, stores, advertisements, other, and family. Five dealt with features, price, brand, style, quality, and size. Three dealt with store visits, number of stores, total visits, and maximum visits to a single store. Finally, one variable dealt with time.

The 18 variables, represented the input for the "nominal approach" to prepurchase style identification. The initial input for the "interval approach" was five variables, one from each area. Number of alternatives, number of information sources, and number of features, were formed by counting endorsements to the nominal variables. Total store visits was used as the interval store variable, and the 4-category time variable was used as an interval variable. The final set of nominal and interval variables is listed in Figure 5.3.
FIGURE 5.3
INFORMATION GATHERING MEASUREMENT
FINAL VARIABLE SET

NOMINAL APPROACH
1. Alternative: Use of Money
2. Alternative: Brands
3. Alternative: Price
4. Source: Friends
5. Source: Salesmen
6. Source: Stores
7. Source: Advertisements
8. Source: Other
9. Source: Family
10. Feature: Price
11. Feature: Brand
12. Feature: Style
13. Feature: Quality
14. Feature: Size
15. Store Visits: Number of Stores
16. Store Visits: Total Visits
17. Store Visits: Maximum Visits to Single Store
18. Time

INTERVAL APPROACH
1. Number of Alternatives Considered
2. Number of Sources Considered
3. Number of Features Considered
4. Number of Store Visits
5. Time
CHAPTER 6

NUMERICAL TAXONOMIC PROCEDURES

PRELIMINARY ANALYSIS

There were three major aspects considered in the evaluation of numerical taxonomic procedures: (1) selection of proximity measures, (2) selection of a clustering algorithm, and (3) selection of a process for evaluating clustering results. Proximity measures and algorithm selections were dealt with by considering the nature of the data, and the exploratory purpose of the research. These selections are discussed in the following two sections. The weakness of descriptive and inferential statistics associated with numerical taxonomy made it necessary to emphasize the evaluation of clustering results. The final section discusses procedures for this purpose.

6.1 PROXIMITY MEASURE SELECTION

The selection of proximity measure is influenced by the nature of the variables. Thus, the discussion considers first the 18-variable nominal approach, and then the 5-variable interval approach.
For the nominal approach two proximity measures were considered - similarity, and city-block distance. Correlation was not considered because of loss of level and scatter mentioned in Chapter 3. The formulation for simple "matching" similarity is given below.

\[ S_{ij} = \sum_{r=1}^{m} \frac{M_{ijr}}{m} \]

where \( S_{ij} \) is the similarity between subjects \( i \) and \( j \); \( M_{ijr} \) is a dichotomous variable with a value of 1 if subjects \( i \) and \( j \) have the same score on variable \( r \) (a value of 0 otherwise); and there are \( m \) variables.

The formulation for city-block distance is given below.

\[ D_{ij}(CB) = \sum_{r=1}^{m} |X_{ri} - X_{rj}| \]

where \( D_{ij}(CB) \) is the city-block distance between subjects \( i \) and \( j \); \( X_{ri} \) is the score of subject \( i \) on variable \( r \); and there are \( m \) variables.

With dichotomous variables "matching" similarity was the complement of city-block distance. Similarity summed the number of matches; city-block distance summed the mismatches. However, these two proximity measures were not equivalent when using multi-category variables. For example a ten-category variable had less impact than the two-category variable using the similarity measure. This was the case since the probability of any pair of subjects "matching" was considerably less with the ten-category variable (0.1
vs 0.5). Thus, the impact of a multi-category variable was swamped by dichotomous variables.

On the other hand the impact of a ten-category variable on city-block distance was greater than the impact of a two-category variable. The between subject difference on a ten-category variable was as much as "+9", but with a dichotomous variable the maximum was "+1". In summing across variables, therefore, the dichotomous variables were swamped by multichotomous variables.

One final comparison of these two proximity measures was considered. Use of city-block with multichotomous variables required stronger assumptions with respect to the relationship among categories. With city-block distance the difference between scores of "1" and "2" is less than between "1" and "3". This was not the case with "matching" similarity.

For this research city-block distance was selected. Two of the 18 variables had more than two categories, time spent and total visits. As indicated in Chapter 5 total store visits was highly correlated with several of the other variables and, therefore, was a candidate for early removal from the set. Thus, the proximity measure decision was based on the impact of the time variable. Time was judged to be a relatively important variable and thus, city-block distance was
chosen in order to increase its impact. A second factor considered when comparing the two proximity measures was differences in assumptions with respect to relationships among categories. With matching similarity the "days" category was assumed equally dissimilar to the "weeks", "months", and "years" categories. With city-block distance "days" to "weeks" was assumed one half the distance from "days" to "months", and one third the distance from "days" to "years". The latter assumption was judged to be a better indication of proximity.

The proximity measure selected for use with the interval approach was the euclidean distance measure shown below.

\[ D_{ij} = \left[ \sum_{r=1}^{m} (X_{ri} - X_{rj})^2 \right]^{\frac{1}{2}} \]

where \( D_{ij} \) is the euclidean distance between subjects \( i \) and \( j \); \( X_{ri} \) is the score of subject \( i \) on variable \( r \); and there are \( m \) variables.

A major alternative proximity measure available for use with the interval variables was correlation. Since this approach standardized scores eliminating the influence of differences in level and scatter, it was not consistent with the intent of this research. For example, a subject whose search variable scores were all high could correlate highly with one whose scores were all low. This would suggest that these
two subjects had similar prepurchase activity patterns. Clearly, a major interest was to differentiate subjects with respect to level and scatter; thus correlation was discarded.

To summarize, city-block distance was used as the proximity measure for the nominal approach. This selection was based on the increased impact of the "time" variable resulting from the use of this measure. The proximity measure used with the interval approach was euclidean distance. It was selected over correlation in order to retain the influence of level and scatter of subjects' scores.

6.2 CLUSTERING ALGORITHM SELECTION

The clustering algorithm selected was Johnson's "diameter" method (Johnson, 1967). This hierarchical clustering scheme sequentially reduced the number of clusters from one cluster per subject to all subjects in one cluster. The cluster characteristic (or criterion) used to determine cluster membership was: the distance between any two subjects in a particular cluster must be less than or equal to the "current clustering distance". As the "current clustering distance" was increased, more and more subjects joined clusters, and small clusters combined to form larger ones. The final step is reached when the "current
clustering distance" reached the maximum distance found in the proximity matrix. At that point all subjects were in one cluster.

Selection of the diameter algorithm was based on three considerations. First, this method required only a rank order assumption of the proximity matrix. This assumption seemed particularly appropriate for the matrix developed using the nominal variables. Second, since subjects joining a cluster must be within the current criterion level from every other cluster member, the clustering criterion resulted in relatively compact clusters. That is, the diameter method led to cluster membership with "likeness" across all members. Since the interest of this research was to identify groups with similar prepurchase styles, the "compactness" associated with the diameter algorithm was desirable.

The final factor of importance in the selection of the diameter method was the hierarchical nature of the algorithm. Because of the exploratory nature of the research, it was desirable to be able to examine clusters at various levels of generality. For example, in the comparison across split sample analyses, were there consistent minor styles; were there consistent major styles; do particular minor styles combine to form particular major styles? (The intended meaning of minor style was a small cluster forming early in the
hierarchical clustering, and a major style was a large cluster forming late in the scheme).

In summary three characteristics of Johnson's diameter method made it appropriate for prepurchase style identification. It was nonmetric, which was consistent with the data. It led to compact clusters, which was consistent with the prepurchase style concept. Finally, it was hierarchical, which was consistent with the exploratory nature of style identification research.

6.3 EVALUATION OF CLUSTERING RESULTS

There were three areas of analysis required with respect to evaluating clustering results. First, Johnson (1967) warned that matrices with identical entries (ties) could result in ambiguous clusters. The importance of this warning was assessed and a procedure was developed to minimize the effect of ties. Second, two procedures for determining the "appropriate" number of clusters were assessed with a view to developing new procedures in this area. Third, a procedure for assessing cluster distinctiveness and cluster replication was developed. Distinctiveness was intended to mean the degree to which clusters differ. A cluster that was separate from other clusters on all dimensions was considered distinctive. A cluster that was similar
to other clusters on some dimensions, and different on others was less distinctive. Replication was intended to mean the degree to which clusters from two different clusterings were similar (or match). Good replication was the case where two clusters from different clusterings were similar on all dimensions.

6.31 Clustering a Proximity Matrix Containing Ties

A simple example is used to illustrate the effect of ties on the clusters identified using Johnson's diameter algorithm. After demonstrating the possibility of ambiguous results when ties are present, assessment is made of the probable occurrence of ties in proximity matrices developed in this research. Finally, it is demonstrated that by reversing subject-order in the proximity matrix and re-clustering, subjects influenced by ties can be identified.

The effect of ties when using the diameter algorithm is illustrated by considering the three subjects plotted below.

```
    .1
   /   
/     
.2     
   /     
    .3
```

The euclidean distance matrix for the three subjects would be:

```
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.41</td>
<td>1.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
```

The clustering criterion is: subjects joining (or forming) a cluster must be a distance less than or equal to the "current criterion level" from all subjects in the cluster. As the current criterion level is reduced sequentially from a distance of "0" to the maximum distance in the matrix, the number of clusters goes from three to one. Clustering the above matrix begins with a search for the minimum non-zero distance. Thus, the first criterion level is 1.00. The question is, which pair clusters first 1-2 or 2-3? The algorithm is such that the search for minimum distance is accomplished by searching the lower left triangle of the distance matrix, beginning at the upper left entry, and checking across each row \((D_{21}', D_{31}', D_{32}')\). Each entry is checked, and the minimum entry encountered is saved. When a distance entry is encountered that equals the one being saved (a tie), the entry found second is ignored. In the example matrix \(D_{21}'\) is encountered before \(D_{32}'\). Thus, the first cluster formed is 1-2.

After the formation of this cluster the distance matrix
is transformed as shown below.

\[
\begin{array}{ccc}
1.2 & 3 \\
1.2 & 0.00 \\
3 & 1.41 & 0.00
\end{array}
\]

The distance between 1-2 and 3 in transformed matrix is the maximum of \( D_{13} \) (1.41) and \( D_{23} \) (1.00). This insures that 3 will not cluster with 1-2 until the criterion level is 1.41. If 1.00 had been used, 3 would cluster with 1-2 at 1.00, thereby violating the clustering criterion. That is, the distance between 3 and 1 would be greater than the current criterion level.

The second criterion level is 1.41 at which point the three subjects are in one cluster. The hierarchical clustering output is:

\[
\begin{array}{ccc}
1 & 2 & 3 \\
1.00 & x & x & x \\
1.41 & x & x & x & x
\end{array}
\]

Although there is no "correct" first-link, the clustering output does not alert the user to this condition. The next step in assessing this problem is to determine under what conditions ties occur.

The number of ties in a matrix is a function of three general factors: the number of subjects, the maximum distance, and the number of significant figures appropriate for the distance measure. For
example, with 144 subjects (the number in Group A
furniture subsample) there are 10,196 inter-subject
distances\(^1\). If distance is based on five variables
standardized to 0.00 - 1.00, the maximum intersubject
distance is 2.24\(^2\). Thus, with three significant
figures there are a \textit{maximum} of 224 distinct distances.
As a result there is a high proportion of ties.

In order to reduce the possibility of ambiguous
results, research attention was directed to the develop-
ment of a procedure for dealing with the ties problem.
It was noticed that if the subject order in the distance
matrix was reversed, the cluster membership changed.
In the 3-stimuli example just discussed the first link
is 3-2 rather than 1-2. The "reverse matrix" clustering
sequence is shown below.

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3.2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.00</td>
<td>3.2</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
<td>0.00</td>
<td>1</td>
<td>1.41</td>
</tr>
<tr>
<td>1</td>
<td>1.41</td>
<td>1.00</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

The result of the reverse matrix clustering on
a larger problem is shown in Figure 6.1. Reversing the
subject-order has the effect of identifying subjects
that are border-line. That is, it identifies subjects
that are between two clusters. In Figure 6.1 subjects

\(^1\) \(n(n-1)/2\)
\(^2\) \(\sqrt[4]{5(1-0)^2}\)
**FIGURE 6.1**
REVERSE MATRIX CLUSTERING

**FORWARD CLUSTERING**

```
  1 1 1 1 1 1 1 1 1 1 1 2
 5 2 6 7 1 3 4 8 9 0 1 2 3 7 6 5 9 8 0

0.10  .  xxx  .  xxx  xxx  xxx  .  xxx  .  xxx  xxx  xxx  xxx  F1
0.14  .  xxxxx  xxxxx  xxx  xxx  xxxxx  xxxxx  xxx  xxx  xxxxx  xxxxx  F2
0.20  .  xxxxxxxx  xxxxxx  xxx  xxxxx  xxxxxxxx  xxxxxxxx  xxxxx  xxxxx  xxxxxxxx  xxxxxxxx  F3
0.28  .  xxxxxxxxxxxx  xxxxxxxxxxx  xxxxxxxx  xxxxxxxxxx  xxxxxxxxxx  xxxxxxxxxx  xxxxxxxxxx  xxxxxxxxxx  F4
0.40  .  xxxxxxxxxxxxxxxxxxxx  xxxxxxxxxxxxxxxxxxxx  xxxxxxxxxxxxxxxxxxxx  xxxxxxxxxxxxxxxxxxxx  F5
0.50  .  xxxxxxxxxxxxxxxxxxxxxxxxxxxx  xxxxxxxxxxxxxxxxxxxxxxxxxxxx  xxxxxxxxxxxxxxxxxxxxxxxxxxxx  F6
0.90  .  xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx  xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx  F7
```

**REVERSE CLUSTERING**

```
  1 1 1 1 2 1 1 1 1 1 1
  6 9 5 4 0 8 7 3 2 1 0 9 7 8 4 5 6 2 3 1

0.10  .  xxx  .  xxx  xxx  xxx  .  xxx  .  xxx  xxx  xxx  xxx  R1
0.14  .  xxxxx  xxxxx  xxx  xxx  xxxxxx  xxxxxx  xxxxx  xxxxx  xxxxxx  xxxxxx  R2
0.20  .  xxxxxxxx  xxxxxxx  xxx  xxxxx  xxxxxxxx  xxxxxxxx  xxxxxxxx  xxxxxxxx  xxxxxxxx  R3
0.28  .  xxxxxxxxxxxx  xxxxxxxxxxx  xxxxxxxx  xxxxxxxxxx  xxxxxxxxxx  xxxxxxxxxx  xxxxxxxxxx  xxxxxxxxxx  R4
0.40  .  xxxxxxxxxxxxxxxxxxxx  xxxxxxxxxxxxxxx  xxxxxxxxxxxxxxx  xxxxxxxxxxxxxxx  xxxxxxxxxxxxxxx  xxxxxxxxxxxxxxx  xxxxxxxxxxxxxxx  xxxxxxxxxxxxxxx  xxxxxxxxxxxxxxx  xxxxxxxxxxxxxxx  R5
0.50  .  xxxxxxxxxxxxxxxxxxxxxxxxx  xxxxxxxxxxxxxxxxxxxxxxxxxxxxx  xxxxxxxxxxxxxxxxxxxxxxxxxxxxx  xxxxxxxxxxxxxxxxxxxxxxxxxxxxx  xxxxxxxxxxxxxxxxxxxxxxxxxxxxx  R6
0.90  .  xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx  xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx  R7
```

*a Clustering level number.*
10 and 11 are directly between the two major clusters. In forward clustering, normal subject-order, 10 and 11 cluster with the major cluster on the right. In reverse clustering, reversed subject-order, 10 and 11 cluster with the major cluster on the left. The term transient is used to describe a subject that changes clusters in this way. Thus, a factor that was not readily apparent at the outset of the analysis of the "ties" problem becomes apparent. A subject whose cluster membership is influenced by ties is a subject that does not fit neatly into one particular cluster.

One final aspect of reverse matrix clustering required consideration. In Figure 6.1 transient subjects were identified by circling sequential clusters in contour fashion, and noting contour inconsistencies between forward and reverse clusterings. This approach was not workable when the number of subjects was large, or when the number of dimensions involved was greater than two. A system for identifying transient subjects was developed. This system identified each subject based on his cluster membership at various clustering levels in the forward matrix hierarchy. These identifiers were then used to check systematically cluster membership in the reverse matrix hierarchy. A description of this procedure is provided in Appendix A.
Reverse matrix clustering was developed to identify transient subjects, but how should transients be handled? The answer to this question was largely a function of the research purpose. Although in some research the interest was categorization of every subject, in this research the interest was identification of major concentrations in the subject-scatter. This was considered consistent with market segmentation analysis. For example, one marketing strategy could be designed to attract buyers interested in one set of factors, and another strategy for buyers interested in another set. This would be done with full recognition that all buyers do not fall neatly into one segment or another. Since it is impractical to design a marketing strategy to match every set of interests, the hope is that the marketing strategies used are sufficiently general to attract "border-line" buyers.

The objective in this research was similar to the market segmentation philosophy. At this stage in identification of prepurchase styles it seemed reasonable to concentrate on major styles. Thus, the clusters given detailed consideration did not include transients.

6.32 Procedures for Assessing the "Appropriate" Number of Clusters

A major consideration in evaluating clustering results was to determine the "appropriate" number of
clusters. In particular using hierarchical clustering, a decision had to be made as to which of the clusters identified in the hierarchical scheme were the most meaningful. The assessment of cluster meaningfulness was based on two factors. First, were the subjects within clusters similar and subjects in different clusters dissimilar, that is, high intra-cluster similarity and low inter-cluster similarity? Second, how distinctive were the clusters? Were clusters clearly different on all dimensions, or only on some of the dimensions? Three procedures provided a guide for determining the appropriate number of clusters.

Joyce and Charron (1966) discussed one approach to this problem - the comparison of average distance within clusters and the average distance between clusters. Conceptually, average within distance, represented the average cluster size, and average between distance represented the average distance between clusters. Joyce and Charron suggested plotting between and within at every criterion level, and selecting clusters where average between distance was great relative to average within. Application of this procedure to the clustering shown in Figure 6.1 is presented below.
This diagram shows that at criterion level 4 (where there are five clusters) the average between/within is 2.25/1.50; at criterion level 5 (three clusters) between/within is 3.50/1.90; and at criterion level 6 (two clusters) between/within is 5.75/2.10. Thus, there appears to be meaningful clusters at either criterion level 5 or 6.

The "between/within" procedure as recommended by Joyce and Charron and as described here, emphasized distances averaged across all clusters. That is, average distance within was the average within all clusters, and average distance between was between all clusters. The same procedure was used to consider whether a particular set of small clusters was sufficiently different to be left as small clusters, or whether they should be combined to form a larger cluster.
Another procedure considered for the selection of most meaningful clusters was a plot of cluster diameter at each criterion level. Cluster diameter was the maximum distance between any two cluster members; thus, it represented a conservative estimate of cluster size. A sharp increase in cluster diameter from one criterion level to the next suggested that two small distinctive clusters had formed one large general cluster. Application of this procedure to the clustering in Figure 6.1 is shown below.

This diagram indicates a gradual increase in cluster size from criterion level 1 to 6, and a greater jump in cluster size between 6 and 7. This suggests criterion level 6 (two clusters) to be the most appropriate cluster set.
The final procedure considered for selecting the most meaningful clusters, was reverse matrix clustering. It was developed, as part of this research, to assess the influence of proximity matrix ties. The procedure identifies transient subjects, those that could be clustered "correctly" in more than one cluster (transients). This provided a useful guide in identifying meaningful clusters. The more transient cluster members in a particular cluster the less meaningful the cluster. To state this another way, the higher the subject intransiency the more meaningful the cluster. By way of example this procedure is applied to part of the Figure 6.1 stimuli.
Comparison of the percent intransients at various criterion levels helps to identify meaningful clusters. In this case there are 66 percent intransient stimuli at criterion level 4, and 100 percent intransient at criterion level 5. This suggests criterion level 5 as a more appropriate cluster set.

The three procedures discussed, average between/within cluster distance plot, cluster diameter plot, and reverse matrix clustering, each provided a guide to the most meaningful cluster set. However, without the benefit of a stimulus plot these procedures do not provide a succinct indication of cluster distinctiveness. The next sub-section discusses a procedure developed to provide an indication of cluster distinctiveness.

6.33 Overlap Assessment

The final step in the discussion of numerical taxonomic procedures is to describe overlap assessment, a procedure developed as part of this research. This procedure helped to assess cluster distinctiveness, and cluster replication; that is, it attempted to provide insight into two questions. First, what was the relative distinctiveness of the clusters identified in a particular cluster analysis? Second, to what extent was there replication of clusters from split sample analyses? Both of these questions were concerned with
the degree of cluster overlap—distinctiveness indicated by a lack of cluster overlap, and replication indicated by a high degree of overlap. Thus, the procedure developed to address these problems was termed overlap assessment.

The value of being able to see clustering results was apparent in the examples discussed earlier. However, with a large sample or greater than three dimensions, plotting subject coordinates was impractical. The problem of plotting a large number of subject coordinates, was dealt with by placing the emphasis on cluster coordinates. The problem of plotting in greater than three dimensions was dealt with by using nonmetric multidimensional scaling to reduce the dimensionality of the cluster-space. Finally, after each cluster location was plotted in the reduced space, the appropriate cluster diameter was plotted around each cluster location. This procedure identified the relative location of each cluster, and provided an indication of cluster size. The degree of overlap in the cluster diameter plots provided an indication of cluster distinctiveness and replication.

To facilitate discussion of overlap assessment in greater detail, it is useful to reconsider the example first presented in Figure 6.1. Assume that after considering between/within distances, a cluster diameter
plot, and analysis of intransients, the "appropriate" number of clusters remained unclear. Assume the clusters under consideration are indicated in Figure 6.2(a). The decision to be made is whether 5 clusters (clusters (1), (2), (7), (4), (5)) or 3 clusters (clusters (3), (7), (6)) are more appropriate. The first step in the application of overlap assessment is to calculate the coordinates for each cluster. This is accomplished by calculating the average across cluster members on each of the variables used to develop the original distance matrix. If 10 variables were used to calculate original distance matrix, the coordinates of a cluster would be a profile of 10 averages. This profile of averages is termed a cluster centroid. In the simple example being considered cluster centroids are the average "x" and "y" for cluster members.

It should be noted that with non-interval data the calculation of cluster centroids is usually inappropriate. With this type of data the mode of cluster members on each characteristic could be used to provide an indication of central tendency. Calculation of cluster centroids based on dichotomous data, however, is appropriate. In this case the centroid represents the proportion of cluster members endorsing each characteristic.

The second step in overlap assessment is to attempt
FIGURE 6.2
OVERLAP ASSESSMENT PROCEDURE: EXAMPLE ONE

FIGURE 6.2 (a)
Major Clusters

FIGURE 6.2 (b)
TORSCA Configuration
Ck: Centroid for Cluster x

FIGURE 6.2 (c)
Diameter Plots
to reduce the dimensionality of the centroid-space. The cluster centroids are based on "m" variables that are usually somewhat intercorrelated. It is useful, therefore, to identify the "k" independent dimensions underlying the "m" variable (where k is less than m). This is done by calculating an inter-centroid euclidean distance matrix, and then analysing this matrix using nonmetric multidimensional scaling.

The objective of nonmetric multidimensional scaling is to locate centroids in space so that - the distance between centroids in the new space have the same rank order as the rank order of the input distance matrix (Kruskel,1964). The centroid arrangement in the new space is referred to as the configuration. A "goodness-of-fit" measure, termed stress, provides an indication of how well the rank order of the configuration distances match the rank order of the input distances. Usually configurations in several dimensionalities are calculated. Selection of the most appropriate dimensionality is based on the goodness-of-fit measure, and the relative ease of interpreting each set of dimensions. The configuration in the selected dimensionality is plotted by using the centroid coordinates provided by the scaling algorithm. It is important to note that the orientation of the configuration axes is relatively arbitrary. Thus, the axes must be rotated to an orientation that is
judged to provide meaningful interpretation.

Figure 6.2(b) shows a plot of centroid locations provided by TORSCA (Young, 1968), a nonmetric multi-dimensional scaling program. The stress in this case is 0.0 indicating a "perfect" fit. This is as expected since in this simple example the final dimensionality is the same as the dimensionality of the input.

There are two cautionary notes with respect to stress. Klar (1969) demonstrated that with few stimuli (less than 10) there was a high probability of a low stress configuration occurring by chance. To avoid this problem the number of stimuli in the centroid-distance matrix should be at least 10. Since there are only seven clusters of interest in the example being discussed, the 20 subjects are also used as centroids.

The second caution concerns the acceptable stress level. The rule of thumb normally used is that a stress of 0.05 is "good", and that 0.10 is "fair" (Kruskal, 1964). In scaling applications where the interest is in assessing the overall configuration, the accurate location of any one stimulus is secondary. Thus, it may be appropriate to accept stress near 0.10 for the sake of lower dimensionality. However, in this application the emphasis is on the specific location of each stimulus, since the amount of overlap is in part a function of the location of the centroid. Thus, stress
of greater than 0.05 is undesirable. Stress greater than this increases the potential error in centroid locations.

The final step in the overlap assessment procedure is to add an estimate of cluster size to the plot of centroid locations. "Cluster diameter" is the clustering criterion in the diameter algorithm. This statistic represents the maximum distance between any two cluster members, and thus, provides a conservative estimate of cluster size. Conceptually, cluster diameter is determined by two cluster members on opposite sides of the cluster scatter. An estimate of cluster size is obtained by drawing a circle of this diameter centered at the cluster centroid. Figure 6.2(c) provides an example.

Tests on known data indicated that this procedure produced very satisfactory results. Statistics other than cluster diameter were considered as estimates of cluster size, for example average within cluster distance. Since cluster diameter was an integral part of the clustering output, and because of its conservative nature, it was selected as the guide to be used in overlap assessment.

Two further considerations with respect to plotting cluster diameter were required. First, when the proximity matrix being clustered was not based on
euclidean distances, cluster diameters were not euclidean. In this case cluster diameter could not be used directly with the scaled centroid plot, since the scaled output was euclidean. For example, in this research city-block distance was used when subject comparisons were based on nominal variables. Thus, cluster diameters were converted from city-block to euclidean prior to plotting. (Comparison of the two distance equations indicated that with dichotomous variables euclidean distance equalled the square root of city-block distance.)

The second consideration before plotting cluster diameters was to match the scale of the cluster diameters to the scale of TORSCA centroid plot. This was done by multiplying each diameter by the scale factor suggested below.

\[
\text{Scale Factor} = \frac{\text{Average Distance, TORSCA Output Distance Matrix}}{\text{Average Distance, Input Distance Matrix}}
\]

The calculation of "Average Distance, Input Distance Matrix" assumes the distances in this matrix are interval. Although this assumption was not required for the TORSCA analysis, it seemed reasonable when using city-block and euclidean distance as proximity measures.

The scaled diameters are plotted in Figure 6.2(c). The congruence between the location of cluster subjects
and cluster location as indicated by the diameter plot suggests that this procedure is a reasonable approach to identifying cluster location and size. The overlap between clusters 1 and 2 (and also 4 and 5) suggests this cluster pair is relatively much less distinctive than cluster 3 compared to cluster 6. If cluster distinctiveness was particularly important this assessment would lead to collapsing clusters 1 and 2 in favor of cluster 3, and clusters 4 and 5 in favor of cluster 6.

This example indicates the use of overlap assessment when the clusters of interest are from a single analysis, that is, as a guide to cluster distinctiveness. Figure 6.3 shows a hypothetical example of the use of overlap assessment when the clusters of interest are from both halves of split-sample analyses. Thus, the emphasis is on distinctiveness and replication. In this case the centroid-distance matrix used as input to the scaling program contains centroids from two different cluster analyses, Group A and Group B. Consideration of clusters from Group A suggests collapsing A2 and A3 due to their relative lack of distinctiveness. Consideration of clusters, from both samples suggests cluster A2-3 is replicated by cluster B1, cluster A4 is replicated by cluster B2, and there are two nonreplicating clusters, A1 and B3.
Clustering identified in the analysis of subgroup "A" of a split sample replication.

○ Clusters identified in the analysis of subgroup "B" of a split sample replication.
To summarize, overlap assessment was developed for the analysis of cluster distinctiveness and cluster replication. The steps involved were: (1) selection of clusters of interest; (2) calculation of the centroids for these clusters; (3) calculation of an inter-centroid distance matrix; (4) analysis of the distance matrix via nonmetric multidimensional scaling; (5) plotting the scaled centroid coordinates in the appropriate dimensionality; and finally, (6) plotting the appropriate cluster diameter around each cluster centroid. This procedure provided a guide for assessing the relative distinctiveness of clusters from a single cluster analysis, and also, for assessing the replication of clusters from split-sample analyses.

Other procedures have been used to identify cluster location and size. One procedure (Jancey, 1966) used factor analysis to reduce the descriptor variables to three orthogonal dimensions. An inter-subject correlation matrix was calculated using three factor scores, and the subjects were clustered. The resulting cluster centroids consisted of a 3-variable profile (three average factor scores) that were plotted as an indication of relative cluster location. For each cluster the standard deviations of scores on the three factors were calculated. For each dimension one standard deviation on either side of the cluster centroid was used as an
indication of cluster size.

Another procedure (Johnson, R.M., 1971) clustered subjects based on a set of desired product attributes; thus cluster centroids were profiles of desired attributes. Multiple discriminant analysis was then used to reduce the attribute-space to two dimensions. An indication of relative cluster location was obtained by plotting the cluster centroids in the reduced space. Cluster size was indicated by plotting circles at each centroid. In this case the circles were used to indicate the number of subjects in the cluster, rather than the physical compactness of the cluster.

For this dissertation low correlations among descriptor variables made preliminary factor analysis inappropriate. Clustering was based on unscaled data, and thus, cluster centroids were multi-variable profiles. Overlap assessment used nonmetric multidimensional scaling to reduce the dimensionality of the centroid space. The reduced-space configuration provided an indication of relative cluster location. Overlap assessment used cluster diameter as an indication of cluster size. Both this approach and the use of standard deviations (Jancey, 1966) were somewhat arbitrary. However, use of cluster diameter was a more conservative approach, and matched the clustering criterion of the "diameter" clustering algorithm.
The other procedures did not include clusters from split-sample analyses. However, it was a logical extension and provided a useful indication of cluster replication.

6.4 SUMMARY

The analysis of numerical taxonomic procedures was discussed in three sections. First, proximity measures appropriate for this research were selected. City-block distance was chosen for the analysis using nominal variables. The primary reason for this choice was that city-block distance, as compared to a similarity measure, increased the relative weight of the "time" variable, (a variable judged to be particularly important). For the analysis using interval variables euclidean distance was chosen. The primary reason for this choice was that the major alternative, correlation, was not influenced by the level or scatter of subjects' scores (two factors considered important in differentiating prepurchase search styles).

The second section discussed the selection of a clustering algorithm. Johnson's diameter method was chosen for three reasons. First, it was non-metric which was particularly appropriate for the analysis based on nominal variables. Second, the clustering criterion used in this algorithm stressed the
identification of compact clusters, which was consistent with the interest in the "major styles" concept. Finally the diameter method was hierarchical and therefore identified clusters of varying degrees of generality. This was consistent with the exploratory nature of the research.

The final section of this chapter discussed the selection and development of procedures for evaluating clustering results. Three procedures were suggested as means for identifying the "appropriate" number of clusters - plotting cluster diameters, plotting between/within cluster distances, and consideration of percent intransients. The procedure for identifying "percent intransients" was developed as part of this research, and was termed reverse matrix clustering. Although the three procedures mentioned above provided a guide in the identification of meaningful clusters, they did not provide a clear indication of cluster distinctiveness. Overlap assessment was developed to address this problem. This procedure was also useful in assessing split-sample cluster replication.
CHAPTER 7

NUMERICAL TAXONOMIC ANALYSIS OF
PREPURCHASE INFORMATION GATHERING

The numerical taxonomic analysis of prepurchase information gathering was done in two segments, an exploration analysis and a re-test analysis. The purpose of the exploration analysis was to determine the "best" approach(s) to prepurchase style identification, and to identify major prepurchase styles. The various approaches tested the effect of nature of measurement, and variables included. This segment of analysis was based on a sample of 287 furniture buyers. The re-test analysis applied the "best" approaches to a second data base. The purpose was both to re-evaluate the approaches, and assess the consistency of prepurchase styles across samples. This segment of the analysis was based on a sample of 259 appliance buyers.

Each approach was analysed using a standardized analysis process. This process is described next.

7.1 STANDARD ANALYSIS PROCESS

Split-sample comparisons were used for both the exploratory and re-test analyses. The terms used to identify the two subsamples were Group A and Group B.
The analysis process consisted of three steps. First, each subsample was clustered using Johnson's diameter algorithm. Second, the clustering results were evaluated to identify clusters of interest in each subsample—these were termed assessment clusters, and were analysed using overlap assessment. A detailed technical description of the analysis process is provided in Appendix B.

For each approach a set of three figures is presented. The first indicates the Group A clusters identified using reverse matrix clustering; the second, the Group B clusters. The third presents the overlap assessment plot. The statistics describing each cluster in the first two figures are presented in a standard form. An example and description of this format is given below.

Example: CLUSTER A1, Figure 7.1

\[
\begin{array}{l}
\text{GR 110: } 39/31 \text{ (80\%)} \\
\text{B/W: } 6.9/5.7 \text{ (1.2)} \\
\text{Diam: } 15.0
\end{array}
\]

The first piece of information, "GR 110", is a cluster identifier (Group 110). The reverse matrix cluster identification system is discussed in Appendix A. The set of figures on the first line, "39/31 (80\%)", is the number of subjects in the cluster: 39 in forward matrix clustering; 31 in reverse matrix clustering; and 80 percent intransience (31/39 \times 100). The second line deals with average distances: "B/W" stands for between/
within; 6.9 is the average distance between clusters prior to the formation of GR 110; and 5.7 is the average distance within clusters prior to the formation of GR 110. That is, Figure 7.1 indicates GR 111 and GR 112 combining to form GR 110. In this case the average city-block distance between subjects in GR 111 and GR 112 is 6.9, and the average within these two clusters is 5.7. The difference (6.9-5.7) is "(1.2)". The final statistic presented is the diameter, "Diam: 15.0" - the maximum distance between any two subjects in GR 110 is 15 city-block units.

With the nominal variables, where city-block distance was used, the units of both averages and diameter were essentially "dissimilar characteristics". With the interval variables each variable was standardized to mean of zero and unit variance prior to the inter-subject euclidean distance calculation. Thus, in this case the distances were "standard deviations".

7.2 EXPLORATION ANALYSIS: NOMINAL VARIABLES

Prior to clustering based on the nominal variables two factors were considered. First, the 18 variables were assessed for redundancy. The phi correlation matrix is presented in Table 7.1. Fourteen intercorrelations showed consistent significant correlations across groups. Eight were labelled "store effect": information from
### TABLE 7.1

**SIGNIFICANT PHI CORRELATIONS**

18 NOMINAL VARIABLES

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A-Save or Spend</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>A-Brand</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>A-Price</td>
<td>0.17</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>I-Friends</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>I-Salesman</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>I-Stores</td>
<td>0.18</td>
<td>0.22</td>
<td>-0.19</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>I-Ads</td>
<td>-0.17</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>I-Other</td>
<td></td>
<td>0.19</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>I-Family</td>
<td>-0.17</td>
<td>0.23</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>F-Price</td>
<td>0.18</td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>F-Brand</td>
<td>-0.43</td>
<td>-0.24</td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>F-Style</td>
<td>0.18</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>F-Quality</td>
<td>-0.17</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>F-Size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>No. Stores</td>
<td>0.27</td>
<td>0.49</td>
<td>0.21</td>
<td>0.17</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Total Visits</td>
<td>0.27</td>
<td>0.51</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Max Visits</td>
<td>0.30</td>
<td>0.48</td>
<td>0.17</td>
<td>0.24</td>
<td>0.29</td>
<td>0.17</td>
<td>0.81</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Time</td>
<td>-0.22</td>
<td>-0.18</td>
<td>-0.17</td>
<td>0.16</td>
<td>-0.39</td>
<td>0.31</td>
<td>-0.19</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

\( N \) of 143, \( r \) of 0.165 is significantly different from 0 at \( \alpha \) of 0.05

When the correlation includes either Total Visits, or Time, product moment correlation was used as a guide.

Time is coded 1,2,3,4 for years, months, weeks and days.
stores, number of stores visited, total visits, and maximum visits to a single store. Total visits had the highest correlations (0.81). Since every variable was considered to contain two aspects of variance, common and unique, the high level of correlation between total visits and other variables indicated that this variable contained little unique variance. In other words, it contained little information not already contained in other variables, and thus, was a candidate for early removal from the nominal set. Apart from total visits variable, the correlations were all relatively low. The highest common variance was 25 percent between information from stores and number of stores visited.

The second consideration was the assumption, implicit to this point, all 287 furniture buyers could be lumped into one group for purposes of style identification. If a particular group was clearly different, it would have been useful to remove it, simplifying further analysis. Sample subgroups based on purchase date, new or used, and product category are compared in Table 7.2. The only subgroups showing consistent significant differences were new vs. used - fewer new purchasers considered information from friends, and more new purchasers considered quality as a feature. The differences were small and not consistent across all factors; thus, were not expected to result in clearly distinctive information
### Table 7.2

**Sub-group Comparisons: 18 Nominal Variables**

<table>
<thead>
<tr>
<th>Purchase Date</th>
<th>A-Save Spend</th>
<th>A-Brand</th>
<th>A-Price</th>
<th>I-Friends</th>
<th>I-Costume</th>
<th>I-Stores</th>
<th>I-Also</th>
<th>I-Other</th>
<th>I-Family</th>
<th>F-Price</th>
<th>F-Brand</th>
<th>F-Style</th>
<th>F-Quality</th>
<th>#-Stores</th>
<th>Total Visits</th>
<th>Max. Visits</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>POST JUNE '69</strong></td>
<td>(41)</td>
<td>0.39</td>
<td>0.73</td>
<td>0.34</td>
<td>0.24</td>
<td>0.36</td>
<td>0.65</td>
<td>0.46</td>
<td>0.17</td>
<td>0.39</td>
<td>0.60</td>
<td>0.07</td>
<td>0.73</td>
<td>0.46</td>
<td>0.43</td>
<td>0.48</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>(31)</td>
<td>0.32</td>
<td>0.77</td>
<td>0.41</td>
<td>0.29</td>
<td>0.41</td>
<td>0.29</td>
<td>0.25</td>
<td>0.19</td>
<td>0.45</td>
<td>0.54</td>
<td>0.03</td>
<td>0.70</td>
<td>0.41</td>
<td>0.48</td>
<td>0.51</td>
<td>1.03</td>
</tr>
<tr>
<td><strong>PRE JUNE '69</strong></td>
<td>(103)</td>
<td>0.29</td>
<td>0.80</td>
<td>0.40</td>
<td>0.16</td>
<td>0.36</td>
<td>0.58</td>
<td>0.32</td>
<td>0.11</td>
<td>0.33</td>
<td>0.58</td>
<td>0.03</td>
<td>0.74</td>
<td>0.45</td>
<td>0.52</td>
<td>0.62</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>(112)</td>
<td>0.31</td>
<td>0.81</td>
<td>0.41</td>
<td>0.12</td>
<td>0.39</td>
<td>0.59</td>
<td>0.27</td>
<td>0.22</td>
<td>0.35</td>
<td>0.57</td>
<td>0.00</td>
<td>0.66</td>
<td>0.53</td>
<td>0.41</td>
<td>0.65</td>
<td>1.06</td>
</tr>
</tbody>
</table>

**New-Used**

| USED | (18) | 0.33 | 0.88 | 0.55 | 0.44 | 0.16 | 0.55 | 0.38 | 0.11 | 0.38 | 0.83 | 0.05 | 0.55 | 0.27 | 0.38 | 0.55 | 0.94 | 0.38 | 2.11 |
|      | (18) | 0.31 | 0.88 | 0.27 | 0.33 | 0.27 | 0.27 | 0.38 | 0.05 | 0.27 | 0.44 | 0.00 | 0.55 | 0.27 | 0.27 | 0.38 | 0.55 | 0.16 | 3.00 |

**NEW**

| (126) | 0.31 | 0.76 | 0.36 | 0.15 | 0.39 | 0.61 | 0.35 | 0.13 | 0.34 | 0.55 | 0.04 | 0.76 | 0.40 | 0.51 | 0.58 | 0.91 | 0.40 | 2.09 |
| (125) | 0.34 | 0.80 | 0.43 | 0.13 | 0.41 | 0.56 | 0.25 | 0.24 | 0.39 | 0.58 | 0.01 | 0.69 | 0.54 | 0.45 | 0.65 | 1.12 | 0.48 | 2.28 |

**Product**

| LIVING ROOM SUITE | (63) | 0.34 | 0.80 | 0.44 | 0.20 | 0.44 | 0.55 | 0.34 | 0.07 | 0.39 | 0.60 | 0.03 | 0.74 | 0.50 | 0.50 | 0.61 | 1.00 | 0.46 | 2.06 |
| UPHOLSTERED CHAIR | (59) | 0.22 | 0.83 | 0.37 | 0.15 | 0.42 | 0.45 | 0.22 | 0.23 | 0.32 | 0.69 | 0.01 | 0.74 | 0.47 | 0.44 | 0.67 | 1.11 | 0.45 | 2.45 |
| DINING ROOM SUITE | (32) | 0.27 | 0.75 | 0.37 | 0.25 | 0.34 | 0.65 | 0.25 | 0.18 | 0.28 | 0.53 | 0.03 | 0.81 | 0.43 | 0.43 | 0.53 | 0.87 | 0.40 | 2.40 |
| BATHROOM SUITE | (34) | 0.44 | 0.61 | 0.33 | 0.27 | 0.55 | 0.33 | 0.16 | 0.38 | 0.50 | 0.05 | 0.72 | 0.55 | 0.55 | 0.55 | 1.11 | 0.61 | 2.11 |
| GROUP A AVERAGE | (40) | 0.23 | 0.70 | 0.35 | 0.05 | 0.26 | 0.61 | 0.41 | 0.14 | 0.35 | 0.64 | 0.11 | 0.67 | 0.38 | 0.47 | 0.47 | 0.73 | 0.35 | 1.94 |
| STANDARD DEVIATION | (34) | 0.40 | 0.53 | 0.12 | 0.50 | 0.68 | 0.31 | 0.12 | 0.50 | 0.53 | 0.00 | 0.59 | 0.50 | 0.50 | 0.78 | 1.31 | 0.50 | 2.10 |
| GROUP B AVERAGE | 0.31 | 0.76 | 0.38 | 0.18 | 0.26 | 0.60 | 0.36 | 0.13 | 0.35 | 0.59 | 0.04 | 0.74 | 0.45 | 0.50 | 0.58 | 0.91 | 0.40 | 2.09 |
| STANDARD DEVIATION | 0.31 | 0.80 | 0.41 | 0.16 | 0.39 | 0.53 | 0.27 | 0.21 | 0.37 | 0.58 | 0.01 | 0.67 | 0.51 | 0.43 | 0.62 | 1.05 | 0.44 | 2.37 |

- **a** Number in each subgroup: top row is Group A, bottom row, Group B
- **b** Significant α of 0.05
- **c** Significant α of 0.10
- **d** The survey was conducted in late fall 1969
- **e** Time is coded 1,2,3,4 for years, months, weeks, and days
gathering styles. However, in keeping with the exploratory nature of the research the "used" furniture buyers (36 in total), were removed for the final two nominal approaches.

There were four clustering approaches using nominal variables. The first clustering was based on all 18 variables, and included all subjects. The second was based on 14 variables, and included all subjects. For this analysis "total visits" and the three "alternatives" variables were removed. The third was based on 12 variables, and only included "new" purchasers. For this analysis "total visits" and the five "features" variables were removed. The final iteration was based on 9 variables, and "new" purchasers. For this analysis "total visits", three "alternatives", and five "features" variables were removed.

The variables included in each nominal approach are indicated in Table 7.3. For each approach the first step in the analysis was the calculation of an inter-subject proximity matrix based on the appropriate variables. As discussed in Chapter 6 with the nominal variables city-block distance was the proximity measure used.

7.21 18-Variable: Figures 7.1, 7.2, 7.3

Reverse matrix clustering of Group A led to the identification of four clusters with high intransiency.
### TABLE 7.3
EXPLORATION ANALYSIS
NOMINAL APPROACHES

<table>
<thead>
<tr>
<th>Alternatives</th>
<th>18-Variables</th>
<th>14-Variables</th>
<th>12-Variables</th>
<th>9-Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Save/Spend</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
</tr>
<tr>
<td>2 Brands</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
</tr>
<tr>
<td>3 Price</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
</tr>
<tr>
<td>Sources</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Salesmen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Stores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Advertisements</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Family</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feature</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Price</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
</tr>
<tr>
<td>11 Brand</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
</tr>
<tr>
<td>12 Style</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
</tr>
<tr>
<td>13 Quality</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
</tr>
<tr>
<td>14 Size</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
</tr>
<tr>
<td>Stores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 Number of Stores</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
</tr>
<tr>
<td>16 Total Visits</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 Maximum to One Store</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(287 Furniture buyers)</td>
<td>36</td>
<td>&quot;USED&quot;</td>
<td>36</td>
<td>&quot;USED&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OUT</td>
<td>OUT</td>
<td>OUT</td>
</tr>
</tbody>
</table>
The 2-cluster solution (A3 and A4) had 90 percent intransient subjects, the 3-cluster solution (A1, A2 and A4) had 80 percent. Assessment of between/within average distances, and changes in diameter did not reveal any clearly appropriate number of clusters. Further, the most compact cluster (A4) had a diameter of 13.0, and an average intra-cluster distance of approximately 7.0. To review, the units of city-block distance were "dissimilar characteristics". Thus, a cluster diameter of 13.0 indicated that the most dissimilar pair in cluster A4 had only five common characteristics of a possible 18. On average Cluster A4 subjects had 11 common characteristics. Since the diameters of A3 and A4 (the 2-cluster solution) were 16 and 13 respectively, and since the over-all diameter was 17, lack of cluster distinctiveness was evident.

Reverse matrix clustering of Group B led to the identification of two clusters; the two had 94 percent intransience. Once again lack of cluster distinctiveness was evident; B1 and B2 had cluster diameters of 15 and 16, and the total sample diameter was 18.

The overlap assessment indicated that clusters A3 and B2 replicated well, suggesting A1 and A2 be dropped in favor of A3. High within-group overlap of clusters A3, A4, B1 and B2 indicated the relative lack of distinctiveness mentioned above. There was, however,
FIGURE 7.1
CLUSTERING BASED ON 18 NOMINAL VARIABLES
GROUP A

[GR 111 15/(N.R.)]

[GR 112 16/(N.R.)]

8 Not Clustered

[GR 121: 27/11 (41%)]
W: 6.8
Diam: 12.0

[GR 122: 28/11 (39%)]
W: 6.3
Diam: 12.0

A1

GR 110: 39/31 (80%)
B/W: 6.9/5.7 (1.2)
Diam: 15.0

b

GR 100: 94/80 (85%)
B/W: 8.4/7.2 (1.2)
Diam: 16.0

A3

A2

GR 120: 55/34 (62%)
B/W: 7.7/6.5 (1.2)
Diam: 14.0

GR A: 144
B/W: 9.7/7.6 (2.1)
Diam: 17.0

A4

GR 210 17/(N.R.)

GR 220 33/(N.R.)

ASSESSMENT CLUSTER PROFILES

NOMINAL VARIABLE NUMBER\(^a\)

<table>
<thead>
<tr>
<th>CLUSTER</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>.58</td>
<td>.80</td>
<td>.45</td>
<td>.12</td>
<td>.19</td>
<td>.87</td>
<td>.54</td>
<td>.35</td>
<td>.32</td>
<td>.64</td>
<td>.03</td>
<td>.74</td>
<td>.29</td>
<td>1.00</td>
<td>1.00</td>
<td>1.64</td>
<td>.70</td>
<td>1.35</td>
</tr>
<tr>
<td>A2</td>
<td>.20</td>
<td>.94</td>
<td>.52</td>
<td>.11</td>
<td>.32</td>
<td>.76</td>
<td>.14</td>
<td>.02</td>
<td>.41</td>
<td>.64</td>
<td>.00</td>
<td>.76</td>
<td>.58</td>
<td>.11</td>
<td>.88</td>
<td>1.14</td>
<td>.29</td>
<td>2.55</td>
</tr>
<tr>
<td>A3</td>
<td>.38</td>
<td>.87</td>
<td>.49</td>
<td>.11</td>
<td>.26</td>
<td>.81</td>
<td>.33</td>
<td>.18</td>
<td>.37</td>
<td>.64</td>
<td>.01</td>
<td>.75</td>
<td>.44</td>
<td>.53</td>
<td>.94</td>
<td>1.38</td>
<td>.49</td>
<td>1.98</td>
</tr>
<tr>
<td>A4</td>
<td>.34</td>
<td>.76</td>
<td>.22</td>
<td>.26</td>
<td>.34</td>
<td>.28</td>
<td>.30</td>
<td>.06</td>
<td>.22</td>
<td>.66</td>
<td>.09</td>
<td>.72</td>
<td>.44</td>
<td>.48</td>
<td>.00</td>
<td>.00</td>
<td>.10</td>
<td>2.24</td>
</tr>
</tbody>
</table>

\(a\) As per Figure 5.3  \(b\) Explanation of cluster statistics provided on page 112  \(c\) (N.R.): no reverse clustering replication
FIGURE 7.2
CLUSTERING BASED ON 18 NOMINAL VARIABLES
GROUP B

GR 110
157 (N.R.)

GR 120
327 (N.R.)

GR 210
587 (N.R.)

GR 220
387 (N.R.)

B1
GR 109: 47/40 (85%) b
B/W: 8.8/5.7 (3.1)
Diam: 15.0

B2
GR 200: 96/95 (99%) b
B/W: 8.1/7.2 (.9)
Diam: 16.0

ASSESSMENT CLUSTER PROFILES

<table>
<thead>
<tr>
<th>NOMINAL VARIABLE NUMBER</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>.20</td>
<td>.77</td>
<td>.15</td>
<td>.20</td>
<td>.27</td>
<td>.15</td>
<td>.22</td>
<td>.25</td>
<td>.20</td>
<td>.40</td>
<td>.00</td>
<td>.60</td>
<td>.40</td>
<td>.35</td>
<td>.00</td>
<td>.00</td>
<td>.05</td>
<td>2.90</td>
</tr>
<tr>
<td>B2</td>
<td>.34</td>
<td>.85</td>
<td>.51</td>
<td>.14</td>
<td>.44</td>
<td>.70</td>
<td>.30</td>
<td>.16</td>
<td>.45</td>
<td>.66</td>
<td>.00</td>
<td>.71</td>
<td>.52</td>
<td>.47</td>
<td>.91</td>
<td>1.49</td>
<td>.60</td>
<td>2.20</td>
</tr>
</tbody>
</table>

a As per Figure 5.3
b Explanation of cluster statistics provided on page 112
FIGURE 7.3

OVERLAP ASSESSMENT: 18 NOMINAL VARIABLES

Stress: 0.036
(3-D: 0.009, 1-D: 0.196)

Group A: Solid circles
Group B: Dashed circles
relatively good between-group replication, indicating two general prepurchase styles. The nature of the styles, as represented by A3-B2 and A4-B1, was suggested by the cluster profiles. The central determinant of the clusterings seemed to be "store effect" with A3-B2 subjects giving considerable attention to information from stores (variable 6), and store visits (variables 15, 16, and 17). The clusters were not differentiated as well on the second dimension. The profiles of clusters A1 and A4 as compared to A2 and B1 suggested the second dimension to be time. A2 and B1 represented less time to make the purchase decision.

7.22 14-Variables: Figures 7.4, 7.5, 7.6

In the second iteration four variables were removed from the 18 variable nominal set, "total visits", and the three "alternatives" variables. Total visits was removed because high correlations with other variables (Table 7.1) indicated considerable measurement redundancy. The "alternatives" variables were removed because they seemed to be different conceptually from the other information gathering variables. In particular the questions measuring features asked for "the main features...considered". This wording attempted to get the respondent to mention features that were of primary importance in the purchase decision. The "alternatives" variables simply asked, "...did you consider...", and
FIGURE 7.4
CLUSTERING BASED ON 14 NOMINAL VARIABLES
GROUP A

A1
GR 110: 12/7 (58%)  
B/Wt: 7.1/4.5 (2.6)  
Diam: 10.0

A2
GR 120: 22/14 (64%)  
B/Wt: 3.1/3.7 (1.4)  
Diam: 8.0

A3
GR 210: 35/26 (74%)  
B/Wt: 5.2/4.3 (0.9)  
Diam: 9.0

GR 220: 26/ (N.R.)

GR 200: 103/87 (84%)  
B/Wt: 6.3/5.4 (0.9)  
Diam: 12.0

ASSESSMENT CLUSTER PROFILES

<table>
<thead>
<tr>
<th>NOMINAL VARIABLE NUMBER</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>.29</td>
<td>.00</td>
<td>.29</td>
<td>.43</td>
<td>.29</td>
<td>.71</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
<td>.43</td>
<td>.57</td>
<td>.43</td>
<td>.57</td>
<td>.43</td>
<td>3.29</td>
</tr>
<tr>
<td>A2</td>
<td>.07</td>
<td>.36</td>
<td>.50</td>
<td>.21</td>
<td>.07</td>
<td>.00</td>
<td>.21</td>
<td>.07</td>
<td>.93</td>
<td>.71</td>
<td>.64</td>
<td>.29</td>
<td>.00</td>
<td>.39</td>
<td>.79</td>
</tr>
<tr>
<td>A3</td>
<td>.15</td>
<td>.30</td>
<td>1.00</td>
<td>.65</td>
<td>.30</td>
<td>.42</td>
<td>.53</td>
<td>.00</td>
<td>.84</td>
<td>.57</td>
<td>.92</td>
<td>1.00</td>
<td>.65</td>
<td>1.30</td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>.32</td>
<td>.32</td>
<td>.20</td>
<td>.24</td>
<td>.06</td>
<td>.12</td>
<td>1.00</td>
<td>.00</td>
<td>.84</td>
<td>.48</td>
<td>.40</td>
<td>.20</td>
<td>.20</td>
<td>1.56</td>
<td></td>
</tr>
</tbody>
</table>

a As per Figure 5.3

b Explanation of cluster statistics provided on page 112
FIGURE 7.5
CLUSTERING BASED ON 14 NOMINAL VARIABLES

GROUP B

GR 110
97 (N.R.)

GR 120: 15/9 (60%)
B/W: 5.7/4.4 (1.3)
Diam: 9.0

GR 100: 48/41 (85%)
B/W: 6.4/4.2 (2.2)
Diam: 11.0

GR 121: 38/29 (76%)
B/W: 5.7/4.2 (1.5)
Diam: 9.0

GR 130: 24/20 (83%)
B/W: 4.3/3.5 (0.8)
Diam: 8.0

GR 210
35 (N.R.)

GR 200: 95/72 (76%)
B/W: 6.5/5.9 (0.6)
Diam: 13.0

GR 220: 60/48 (80%)
B/W: 6.5/5.1 (1.4)
Diam: 12.0

GR 221: 22/18 (82%)
B/W: 5.9/4.1 (1.8)
Diam: 9.0

ASSESSMENT CLUSTER PROFILES

NOMINAL VARIABLE NUMBER

<table>
<thead>
<tr>
<th>CLUSTER</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>.27</td>
<td>.27</td>
<td>.90</td>
<td>.09</td>
<td>.09</td>
<td>.36</td>
<td>1.00</td>
<td>.00</td>
<td>.54</td>
<td>.81</td>
<td>.36</td>
<td>1.00</td>
<td>.63</td>
<td>3.63</td>
</tr>
<tr>
<td>B2</td>
<td>.23</td>
<td>.30</td>
<td>.15</td>
<td>.15</td>
<td>.03</td>
<td>.03</td>
<td>.57</td>
<td>.00</td>
<td>.69</td>
<td>.34</td>
<td>.26</td>
<td>.03</td>
<td>.23</td>
<td>3.38</td>
</tr>
<tr>
<td>B3</td>
<td>.00</td>
<td>.48</td>
<td>.96</td>
<td>.37</td>
<td>.24</td>
<td>.37</td>
<td>.20</td>
<td>.03</td>
<td>.72</td>
<td>.55</td>
<td>.24</td>
<td>.96</td>
<td>.48</td>
<td>2.06</td>
</tr>
<tr>
<td>B4</td>
<td>.22</td>
<td>.11</td>
<td>.55</td>
<td>.22</td>
<td>.22</td>
<td>.33</td>
<td>.94</td>
<td>.00</td>
<td>.83</td>
<td>.27</td>
<td>.83</td>
<td>.72</td>
<td>.83</td>
<td>1.72</td>
</tr>
</tbody>
</table>

\(a\) As per Figure 5.3

\(b\) Explanation of cluster statistics provided on page 112
FIGURE 7.6
OVERLAP ASSESSMENT: 14 NOMINAL VARIABLES

Stress: 0.044
(3-D:0.015,1-D:0.070)

Group A: Solid circles
Group B: Dashed circles
therefore, had the potential of being endorsed even when the alternative was of minor significance to the respondent.

Reverse matrix clustering of Group A identified six clusters with high intransiency. The 2-cluster solution had 76 percent intransient subjects; the 4-cluster solution had 50 percent. The diameters of the two general clusters were 12.0. This indicated that the most dissimilar pair in each cluster only had two common characteristics (14-12) - a lack of compact clusters.

Reverse matrix clustering of Group B identified seven clusters with high intransiency. The 2-cluster solution had 78 percent intransient subjects; the 4-cluster solution had 53 percent. Once again the cluster diameters indicated lack of cluster compactness. The 2-cluster solution had cluster diameters of 11 and 13.

Overlap assessment suggested two major groupings A1, A2, B1, B2, and A3, A4, B3, B4. Thus, 2-cluster solutions seemed appropriate for both Group A and Group B. The within-group overlap indicated lack of distinctive clusters. However, there was good between-groups replication suggesting two general styles. The reduced-space cluster plot required two dimensions to represent cluster locations adequately (stress of 0.04). Consideration of cluster profiles indicated the dimen-
sions to be "time" and "stores". The orientation of the dimensions is shown in Figure 7.6. The two extremes on the "time" dimensions were clusters A2 and A3 (3.79 and 1.30 on variable 18). Clusters A3 and A4 (also B2 and B1) indicated differences on the "stores" dimension. The cluster A3 represented use of information from stores, visiting several stores, and several visits to a single store (variables 6, 15, 17). Cluster A4 represented little use of information from stores, visiting few stores, and few visits to a single store.

The two general styles identified differed primarily in terms of time spent on decision making. Clusters A3/ A4 and B3/B4 spent from "months to years", and clusters A1/A2 and B1/B2 spent from "days to weeks".

7.23 12-Variables: Figures 7.7, 7.8, 7.9

In the third iteration "total visits" and the five "features" variables were removed. Total visits, as in the 14-variable analysis, was removed because of high intercorrelations with other variables. The reason for trying style identification with the "features" variables removed was the wording of the question used to measure this area. The question, although open ended, provided five example features; "What were the main features, for example, size, style, price, sturdiness, store you considered...." The response distribution showed four major responses - price (58 percent of respondents),
FIGURE 7.7
CLUSTERING BASED ON 12 NOMINAL VARIABLES
GROUP A

A1 [GR 100: 7/6 (86%)]
  W: 3.9
  Diam: 5.0

A2 [GR 210: 39/24 (62%)]
  W: 4.6
  Diam: 9.0

A3 [GR 220: 31/29 (94%)]
  B/W: 5.3/3.8 (1.5)
  Diam: 10.0

A4 [GR 310: 24/16 (67%)]
  B/W: 5.4/3.6 (1.8)
  Diam: 8.0

A5 [GR 320: 25/20 (80%)]
  W: 3.4
  Diam: 6.0

ASSESSMENT CLUSTER PROFILES

<table>
<thead>
<tr>
<th>CLUSTER</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>15</th>
<th>17</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>.00</td>
<td>.33</td>
<td>.66</td>
<td>.83</td>
<td>.66</td>
<td>.16</td>
<td>1.00</td>
<td>.00</td>
<td>.66</td>
<td>.00</td>
<td>.50</td>
<td>2.16</td>
</tr>
<tr>
<td>A2</td>
<td>.08</td>
<td>.66</td>
<td>.33</td>
<td>.12</td>
<td>.87</td>
<td>.58</td>
<td>.45</td>
<td>.12</td>
<td>.62</td>
<td>.70</td>
<td>.58</td>
<td>1.83</td>
</tr>
<tr>
<td>A3</td>
<td>.58</td>
<td>.89</td>
<td>.62</td>
<td>.06</td>
<td>.20</td>
<td>.93</td>
<td>.41</td>
<td>.31</td>
<td>.37</td>
<td>1.00</td>
<td>.68</td>
<td>1.37</td>
</tr>
<tr>
<td>A4</td>
<td>.30</td>
<td>.85</td>
<td>.20</td>
<td>.10</td>
<td>.00</td>
<td>.20</td>
<td>.35</td>
<td>.00</td>
<td>.15</td>
<td>.10</td>
<td>.00</td>
<td>1.75</td>
</tr>
<tr>
<td>A5</td>
<td>.25</td>
<td>.70</td>
<td>.05</td>
<td>.10</td>
<td>.25</td>
<td>.35</td>
<td>.15</td>
<td>.10</td>
<td>.05</td>
<td>.30</td>
<td>.10</td>
<td>3.60</td>
</tr>
</tbody>
</table>

a A1 per Figure 5.3
b Explanation of cluster statistics provided on page 112
Figure 7.8
Clustering Based on 12 Nominal Variables
Group B

B2  
GR 221: 13/11 (85%)  
B/W: 4.4/2.7 (1.7)  
Diam: 6.0

B3  
GR 220: 31/(N.R.)

B1  
GR 210: 6/(N.R.)

B4  
GR 310: 16/13 (81%)  
B/W: 4.1/1.9 (2.2)  
Diam: 6.0

B5  
GR 320: 29/20 (69%)  
B/W: 5.0/3.4 (1.6)  
Diam: 8.0

GR 200: 67/41 (61%)  
B/W: 5.6/4.6 (1.0)  
Diam: 10.0

GR B: 125  
B/W: 6.9/5.4 (1.5)  
Diam: 13.0

Assessment Cluster Profiles

<table>
<thead>
<tr>
<th>Cluster</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>15</th>
<th>17</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>.85</td>
<td>1.00</td>
<td>1.00</td>
<td>.00</td>
<td>.57</td>
<td>1.00</td>
<td>1.00</td>
<td>.57</td>
<td>1.00</td>
<td>.85</td>
<td>.85</td>
<td>1.71</td>
</tr>
<tr>
<td>B2</td>
<td>.90</td>
<td>.90</td>
<td>.63</td>
<td>.18</td>
<td>.36</td>
<td>.54</td>
<td>.09</td>
<td>.00</td>
<td>.27</td>
<td>1.00</td>
<td>.45</td>
<td>1.18</td>
</tr>
<tr>
<td>B3</td>
<td>.35</td>
<td>.89</td>
<td>.21</td>
<td>.07</td>
<td>.39</td>
<td>.28</td>
<td>.00</td>
<td>.32</td>
<td>.42</td>
<td>.28</td>
<td>.57</td>
<td>2.03</td>
</tr>
<tr>
<td>B4</td>
<td>.00</td>
<td>1.00</td>
<td>.15</td>
<td>.00</td>
<td>.23</td>
<td>.00</td>
<td>.38</td>
<td>.07</td>
<td>.15</td>
<td>.07</td>
<td>.00</td>
<td>3.53</td>
</tr>
<tr>
<td>B5</td>
<td>.05</td>
<td>.95</td>
<td>.80</td>
<td>.15</td>
<td>.10</td>
<td>.90</td>
<td>.40</td>
<td>.15</td>
<td>.35</td>
<td>1.00</td>
<td>.55</td>
<td>2.90</td>
</tr>
</tbody>
</table>

a As per Figure 5.3
b Explanation of cluster statistics provided on page 112
FIGURE 7.9
OVERLAP ASSESSMENT: 12 NOMINAL VARIABLES

Stress: 0.01
(4-D: 0.009, 2-D: 0.089)

Group A: Solid circles
Group B: Dashed circles
style (64 percent), quality (48 percent), and size (47 percent). The similarity between the major response categories and the example features indicated the possibility of a response bias problem.

Reverse matrix clustering of Group A identified 7 clusters of high intransiency. The 3-cluster solution had 91 percent intransients; the 5-cluster solution had 75 percent. Reverse matrix clustering of Group B also identified 7 clusters of high intransiency. The 3-cluster solution had 66 percent intransients; the 5-cluster solution had 63 percent. As in the earlier analyses there was no clearly "appropriate" set of clusters indicated by the "between/within" and "diameter" statistics.

The nonmetric scaling of the 10 cluster centroids required three dimensions to get a "good fit" between the centroid-distance matrix and the reduced-space configuration. In two dimensions the stress was "fair" (0.09); in three dimensions it was "excellent" (0.01). The distribution of clusters indicated that two dimensions accounted for most of the cluster differences. The two main dimensions were "time" and "stores". The third dimension differentiated cluster A1, a small cluster accounting for only five percent of Group A. Comparison of the cluster A1 profile with the other profiles indicated the third dimension to be associated with two
factors - information from friends, and consideration of alternative brands.

Because of high within-group overlap and lack of between-group replication, no information gathering styles were identified. The overlap assessment plot indicated that the clusters were distributed in a small space relative to cluster size. This suggested a relatively uniform distribution of subjects within the subject-space. A uniform distribution of subjects would explain the lack of cluster replication, since in this case cluster formation would be somewhat arbitrary.

To illustrate the clustering of a uniform distribution the distribution plotted below was analysed.

These stimuli were clustered twice using two different stimulus orders in the proximity matrix. The numbers beside the plots indicate the two proximity matrix stimulus orders. The upper number was used for clustering A, and the lower for clustering B. Reverse matrix
clustering was used for both A and B. At the two-cluster level substantially different clusterings resulted. These are indicated below.

"A" - Clustering

"B" - Clustering

The differences were due to the stimulus order in the proximity matrix. A plot of the four cluster centroids and cluster diameters resulted in non-replicating clusters with a high degree of cluster overlap as indicated below.

This simple example helped to explain the meaning of the 12-variable overlap assessment plot. "Nonreplication and high overlap" suggested a relatively uniform distribution of subjects. In other words there was a lack of
subject concentration(s) in the subject-space.

7.24 9-Variables: Figure 7.10, 7.11, 7.12

The final iteration using nominal variables removed "total visits", the three "alternatives" variables, and the five "features" variables. The potential weaknesses of these variables have been discussed.

Reverse matrix clustering of Group A identified a five-cluster solution with 65 percent intransient subjects. The clustering between the five-cluster solution, and one-cluster solution indicated no consistency between forward and reverse matrix clusterings. This indicated an incomplete hierarchy; intransient clusters at the 5-cluster level, but no intransient clusters at the 4-, 3-, or 2-cluster levels.

A simple two-dimensional example of subject concentrations that result in an incomplete hierarchy is shown below. The three clusters are spaced equi-distant from each other. In this case the clusters indicate high intransiency at the 3-cluster level, but there is no "correct" two cluster solution.

[Diagram of clusters]
FIGURE 7.10
CLUSTERING BASED ON 9 NOMINAL VARIABLES
GROUP A

GR 110: 8/0 (N.R.)

A1 [GR 120: 28/14 (50%)]
   B/W: 3.7/2.4 (1.3)
   Diam: 7.0
   b [GR 100: 36/7 (N.R.)]

A2 [GR 210: 23/17 (74%)]
   B/W: 3.1/2.2 (0.9)
   Diam: 6.0
   [GR 200: 55/7 (N.R.)]

A3 [GR 220: 32/29 (91%)]
   W: 3.17
   Diam: 5.0
   [GR A: 126
   W: 4.4
   Diam: 9.0]

A4 [GR 310: 16/13 (81%)]
   B/W: 3.1/1.9 (1.2)
   Diam: 5.0
   [GR 300: 28/7 (N.R.)]

A5 [GR 320: 12/8 (67%)]
   W: 2.5
   Diam: 5.0
   7 Not Clustered

ASSESSMENT CLUSTER PROFILES

<table>
<thead>
<tr>
<th>NOMINAL VARIABLE NUMBER</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>15</th>
<th>17</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLUSTER</td>
<td>A1</td>
<td>.14</td>
<td>.07</td>
<td>.92</td>
<td>.50</td>
<td>.78</td>
<td>.42</td>
<td>1.00</td>
<td>.71</td>
</tr>
<tr>
<td></td>
<td>A2</td>
<td>.05</td>
<td>.05</td>
<td>.17</td>
<td>.47</td>
<td>.00</td>
<td>.29</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>A3</td>
<td>.03</td>
<td>.72</td>
<td>.93</td>
<td>.10</td>
<td>.00</td>
<td>.45</td>
<td>.72</td>
<td>.48</td>
</tr>
<tr>
<td></td>
<td>A4</td>
<td>.13</td>
<td>.31</td>
<td>.44</td>
<td>.19</td>
<td>.06</td>
<td>.00</td>
<td>.31</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>A5</td>
<td>.12</td>
<td>.00</td>
<td>.12</td>
<td>.37</td>
<td>.00</td>
<td>.00</td>
<td>.62</td>
<td>.87</td>
</tr>
</tbody>
</table>

a As per Figure 5.3
b Explanation of cluster statistics provided on page 112
FIGURE 7.11
CLUSTERING BASED ON 9 NOMINAL VARIABLES
GROUP B

GR 110; 23/(N.R.)  
GR 120; 20/(N.R.)  
GR 210; 25/(N.R.)  
GR 231; 97/(N.R.)  
GR 232: 39/19 (49%)  
\[ \text{B/W: 3.5/2.3 (1.2)} \]  
\[ \text{Diam: 6.0} \]  
GR 100; 43/25 (58%)  
\[ \text{B/W: 4.0/2.8 (1.2)} \]  
\[ \text{Diam: 8.0} \]  
\[ \text{b} \]  
GR B; 125  
\[ \text{B/W: 6.3/3.8 (2.5)} \]  
\[ \text{Diam: 10.0} \]  
\[ \text{GR 200; 82/68 (83%)} \]  
\[ \text{B/W: 4.3/3.5 (0.8)} \]  
\[ \text{Diam: 9.0} \]  

ASSESSMENT CLUSTER PROFILES

<table>
<thead>
<tr>
<th>NOMINAL VARIABLE NUMBERS\textsuperscript{a}</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>15</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLUSTERS \textsuperscript{b}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>.24</td>
<td>.32</td>
<td>.08</td>
<td>.24</td>
<td>.04</td>
<td>.16</td>
<td>.16</td>
<td>.16</td>
<td>.304</td>
</tr>
<tr>
<td>B2</td>
<td>.00</td>
<td>.52</td>
<td>.94</td>
<td>.52</td>
<td>.05</td>
<td>.57</td>
<td>1.00</td>
<td>.31</td>
<td>.78</td>
</tr>
</tbody>
</table>

\textsuperscript{a} As per Figure 5.3

\textsuperscript{b} Explanation of cluster statistics provided on page 112
FIGURE 7.12
OVERLAP ASSESSMENT: 9 NOMINAL VARIABLES

Stress: 0.027
(3-D:0.021,1-D:0.087)

Group A: Solid circles
Group B: Dashed circles
Reverse matrix clustering of Group B identified four clusters of relatively high intransiency. The two-cluster solution had 75 percent intransient subjects. Clustering levels other than the two-cluster level indicated very little reverse matrix replication. Thus, Group A and Group B were quite different. Group A clustered best at the 5-cluster level, and Group B clustered best at the 2-cluster level.

The overlap assessment plot required two dimensions for an "excellent" representation of cluster location (stress 0.03). Comparison of cluster profiles indicated the major dimensions separating clusters were "time" (variable 18), and "stores" (variables 6, 15, 17). The two groups showed considerable within-group overlap and no clear between-group cluster replication. The plot indicated that the seven assessment clusters were grouped in a relatively confined area. As in the 12-variable approach the confined nature of the clusters suggested a uniform distribution of subjects. Since there was no cluster replication, no prepurchase styles were identified.

7.25 Nominal Variable Approaches: Summary

Four analysis approaches were done based on the 18 nominal variables listed in Figure 5.3. The first approach considered all 18 variables. The second removed
"total visits" because of redundancy, and three "alternatives" variables because of question wording. The third removed "total visits" because of redundancy, and five "features" variables because of potential response bias. The final analysis approach removed "total visits", "alternatives", and "features".

A general finding was the presence of two major dimensions of prepurchase information gathering, a "time" dimension, and a "stores" dimension. In the third iteration ("total visits" and "features" removed) description of cluster location required three dimensions. In addition to "time" and "stores", a small cluster (representing 5 percent of Group A) was separated from the other clusters because of higher use of information from friends, and lower consideration of alternative brands.

Major information gathering styles were identified only in the first and second approaches and were of a very general nature. In the first approach two general styles, differing in terms of "stores", replicated across split sample analysis. In the second approach two general styles, differing in terms of "time" and "stores", replicated across split sample analysis. In the third and fourth approaches there was no cluster replication across subsamples. The overlap assessment plot of the final two approaches suggested that a uniform distribution
of subjects had resulted in somewhat arbitrary cluster formation, in which case little cluster replication would be expected.

7.3 EXPLORATION ANALYSIS: INTERVAL VARIABLES

Five variables provided the base for this section of the analysis: number of alternatives considered, number of features considered, number of information sources considered, number of store visits, and time. The product moment correlation matrix for the five interval variables is presented in Table 7.4. Although most of the correlations were statistically significant, the level of the common variance between any two variables was relatively low (the range of pair-wise $R^2$ was from 0.00 to 0.13). This indicated that measurement redundancy was not a concern with this data.

Subgroups based on purchase date, new-used, and product were compared, Table 7.5. There were no consistent significant differences. Thus, all attempts at style identification using interval variables included all 287 furniture purchasers.

There were three clustering approaches based on interval variables. The first included all five variables. In the second "number of alternatives" was removed, and in the third both the "alternatives" and "features" variables were removed. The variables included in each
### TABLE 7.4

PRODUCT MOMENT CORRELATION:
5 INTERVAL VARIABLES

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Alternatives</td>
<td>1.00&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Information Sources</td>
<td>.02&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.22</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Features</td>
<td>.10</td>
<td>.00</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.13</td>
<td>.22</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Visits</td>
<td>.18</td>
<td>.36</td>
<td>.10</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.25</td>
<td>.35</td>
<td>.29</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>5 Time&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-.24</td>
<td>-.18</td>
<td>-.20</td>
<td>-.20</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>-.26</td>
<td>-.25</td>
<td>-.24</td>
<td>-.29</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<sup>a</sup> N of 143, r of <sup>+</sup> 0.165 is significantly different from 0 at α of 0.05

<sup>b</sup> Time is coded 1,2,3,4 for years, months, weeks and days

<sup>c</sup> Upper correlation: Group A, lower correlation: Group B
### TABLE 7.5

**SUB-GROUP COMPARISONS**

5 INTERVAL VARIABLES

<table>
<thead>
<tr>
<th>PURCHASE DATE</th>
<th>No. Alternatives</th>
<th>No. Sources</th>
<th>No. Features</th>
<th>No. Store Visits</th>
<th>Timed</th>
</tr>
</thead>
<tbody>
<tr>
<td>POST JUNE '69</td>
<td>(41)(^a)</td>
<td>1.46</td>
<td>2.29</td>
<td>2.31</td>
<td>4.73</td>
</tr>
<tr>
<td></td>
<td>(31)</td>
<td>1.51</td>
<td>1.90</td>
<td>2.19</td>
<td>4.93</td>
</tr>
<tr>
<td>PRE JUNE '69</td>
<td>(103)</td>
<td>1.50</td>
<td>1.89</td>
<td>2.34</td>
<td>4.44</td>
</tr>
<tr>
<td></td>
<td>(112)</td>
<td>1.53</td>
<td>1.97</td>
<td>2.20</td>
<td>5.20</td>
</tr>
<tr>
<td>NEW-USED</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USED</td>
<td>(18)</td>
<td>1.77</td>
<td>2.05</td>
<td>2.11</td>
<td>5.27</td>
</tr>
<tr>
<td></td>
<td>(18)</td>
<td>1.22</td>
<td>1.61</td>
<td>1.55(^b)</td>
<td>3.11</td>
</tr>
<tr>
<td>NEW</td>
<td>(126)</td>
<td>1.45</td>
<td>2.00</td>
<td>2.37</td>
<td>4.42</td>
</tr>
<tr>
<td></td>
<td>(125)</td>
<td>1.57</td>
<td>2.00</td>
<td>2.29</td>
<td>5.44</td>
</tr>
<tr>
<td>PRODUCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIVING ROOM SUITE</td>
<td>(63)</td>
<td>1.60</td>
<td>2.03</td>
<td>2.39</td>
<td>4.69</td>
</tr>
<tr>
<td></td>
<td>(59)</td>
<td>1.42</td>
<td>1.81</td>
<td>2.37</td>
<td>5.57</td>
</tr>
<tr>
<td>UPHOLSTERED CHAIR</td>
<td>(32)</td>
<td>1.50</td>
<td>1.96</td>
<td>2.25</td>
<td>4.31</td>
</tr>
<tr>
<td></td>
<td>(34)</td>
<td>1.55</td>
<td>1.91</td>
<td>1.88</td>
<td>3.26</td>
</tr>
<tr>
<td>DINING ROOM SUITE</td>
<td>(15)</td>
<td>1.46</td>
<td>2.33</td>
<td>2.40</td>
<td>5.20</td>
</tr>
<tr>
<td></td>
<td>(18)</td>
<td>1.38</td>
<td>2.00</td>
<td>2.38</td>
<td>5.16</td>
</tr>
<tr>
<td>BEDROOM SUITE</td>
<td>(34)</td>
<td>1.29</td>
<td>1.85</td>
<td>2.29</td>
<td>4.11</td>
</tr>
<tr>
<td></td>
<td>(37)</td>
<td>1.78</td>
<td>2.25</td>
<td>2.12</td>
<td>6.34</td>
</tr>
<tr>
<td>TOTAL GROUP STATISTICS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVERAGES</td>
<td></td>
<td>1.49</td>
<td>2.00</td>
<td>2.34</td>
<td>4.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.53</td>
<td>1.95</td>
<td>2.20</td>
<td>5.14</td>
</tr>
<tr>
<td>STANDARD DEVIATIONS</td>
<td></td>
<td>.79</td>
<td>1.23</td>
<td>.88</td>
<td>4.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.85</td>
<td>1.20</td>
<td>.98</td>
<td>4.38</td>
</tr>
</tbody>
</table>

\(^a\) Number in each subgroup: top row is Group A; bottom row, Group B

\(^b\) Significant at \(\alpha\) of 0.05

\(^c\) The survey was conducted in late fall 1969

\(^d\) Time is coded 1,2,3,4 for years, months, weeks and days
approach are indicated in Table 7.6. For each approach the first step in the analysis was the calculation of an inter-subject proximity matrix based on the appropriate variables. This was done by standardizing each variable to zero mean and unit variance, and then calculating the Euclidean distance between each subject pair.

7.3.1 5-Variables: Figures 7.13, 7.14, 7.15

Reverse matrix clustering of Group A indicated 11 clusters with high intransiency. The two-cluster level had 94 percent intransient subjects. The four-cluster level had 40 percent. The nine small clusters had 74 percent. The drop in percent intransients at the intermediate clustering level indicated that although the small clusters consistently ended up in the same general cluster, there were no "correct" intermediate clusters. Consideration of average cluster distances and cluster diameters showed that the two general clusters were each almost as large as the total group cluster. This indicated a lack of cluster distinctiveness at the general cluster level.

Reverse matrix clustering of Group B also indicated 11 clusters with high intransiency. The two-cluster level had 87 percent intransient subjects. The six-cluster level had 81 percent, and the final 7 clusters had 66 percent. The lack of "correct" intermediate
<table>
<thead>
<tr>
<th></th>
<th>5 Variables</th>
<th>4 Variables</th>
<th>3 Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Number of Alternatives</td>
<td>OUT</td>
<td>OUT</td>
</tr>
<tr>
<td>2.</td>
<td>Number of Sources</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Number of Features</td>
<td></td>
<td>OUT</td>
</tr>
<tr>
<td>4.</td>
<td>Number of Store Visits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Time</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample: For each approach, all 287 furniture buyers were included.
clusters found in Group A was not found in Group B. Instead, lack of replicating small clusters was observed. A further difference between the Group A and Group B clusterings was that the former had two general clusters while the latter had three. These inconsistencies between subsamples led to the suspicion that the distribution of subjects in the subject-space was relatively uniform, and thus, cluster membership was arbitrary. This suspicion was supported by the overlap assessment plot.

Nonmetric scaling of the five-variable clusters required three dimensions to represent the centroid-space adequately (the stress in three dimensions was 0.05; in two was 0.08). The dimensions in Figure 7.15 were identified by comparing relative cluster locations with cluster profile scores. For example in the bottom diagram clusters A5, A6, and A7 had few store visits and clusters B1 and A1 had many store visits. The objective when orienting the dimensions was to have the rank order of the clusters on the dimension match the rank order of cluster averages on a particular variable. Although the matching was not perfect, clusters with large averages were toward one end and clusters with small averages toward the other. Two dimensions were identified: number of visits and number of features. The nature of the third dimension was not clear.
FIGURE 7.13
CLUSTERING BASED ON 5 INTERVAL VARIABLES

GROUP A

[5 Not Clustered]

GR 121: 21/21 (100%)  
B/W: 2.1/1.4 (0.7)  
Diam: 2.8

GR 122: 19/16 (85%)  
B/W: 1.9/1.3 (0.4)  
Diam: 2.7

GR 110: 13/11 (85%)  
W: 3.2  
Diam: 3.2

GR 120: 46/46 (100%)  
B/W: 2.4/1.9 (0.5)  
Diam: 3.8

GR 110: 64/62 (97%)  
B/W: 3.9/2.3 (1.5)  
Diam: 6.2

GR A: 144  
B/W: 3.3/2.5  
Diam: 7.3

[1 Not Clustered]

GR 210: 307 (N.R.)

GR 211: 22/15 (68%)  
B/W: 2.5/1.8 (0.7)  
Diam: 3.8

GR 212: 8 (W.R.)

GR 221: 15/13 (87%)  
W: 1.7  
Diam: 3.2

GR 220: 497 (N.R.)

GR 222: 10/10 (100%)  
W: 1.2  
Diam: 2.0

GR 223: 14/10 (71%)  
W: 1.6  
Diam: 3.2

GR 224: 10/10 (100%)  
W: 1.4  
Diam: 2.0

ASSESSMENT CLUSTER PROFILES

<table>
<thead>
<tr>
<th>VARIABLE NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>5(R)</td>
</tr>
<tr>
<td>4(R)</td>
</tr>
<tr>
<td>3(R)</td>
</tr>
<tr>
<td>2(R)</td>
</tr>
<tr>
<td>1(R)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CLUSTER</th>
<th>1(R)</th>
<th>1(S)</th>
<th>2(R)</th>
<th>2(S)</th>
<th>3(R)</th>
<th>3(S)</th>
<th>4(R)</th>
<th>4(S)</th>
<th>5(R)</th>
<th>5(S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>2.00</td>
<td>.64</td>
<td>3.81</td>
<td>1.46</td>
<td>2.72</td>
<td>.43</td>
<td>8.54</td>
<td>.93</td>
<td>2.00</td>
<td>-.10</td>
</tr>
<tr>
<td>A2</td>
<td>2.04</td>
<td>.69</td>
<td>1.91</td>
<td>-.07</td>
<td>2.73</td>
<td>.45</td>
<td>3.76</td>
<td>-.17</td>
<td>1.39</td>
<td>-.73</td>
</tr>
<tr>
<td>A3</td>
<td>.30</td>
<td>-.73</td>
<td>3.31</td>
<td>1.05</td>
<td>1.68</td>
<td>-.74</td>
<td>5.18</td>
<td>.15</td>
<td>2.09</td>
<td>.00</td>
</tr>
<tr>
<td>A4</td>
<td>.80</td>
<td>-.87</td>
<td>1.33</td>
<td>-.54</td>
<td>1.73</td>
<td>-.68</td>
<td>3.06</td>
<td>-.34</td>
<td>1.46</td>
<td>-.65</td>
</tr>
<tr>
<td>A5</td>
<td>1.60</td>
<td>.13</td>
<td>.80</td>
<td>-.97</td>
<td>1.00</td>
<td>-1.51</td>
<td>2.40</td>
<td>-.49</td>
<td>2.60</td>
<td>.52</td>
</tr>
<tr>
<td>A6</td>
<td>1.40</td>
<td>-.11</td>
<td>.90</td>
<td>-.89</td>
<td>2.20</td>
<td>-.15</td>
<td>3.20</td>
<td>-.31</td>
<td>3.70</td>
<td>1.67</td>
</tr>
<tr>
<td>A7</td>
<td>53</td>
<td>-1.21</td>
<td>1.06</td>
<td>-.75</td>
<td>3.00</td>
<td>.74</td>
<td>1.46</td>
<td>-.71</td>
<td>2.93</td>
<td>.87</td>
</tr>
</tbody>
</table>

a As per Figure 5.3
b (R), raw average; (S), standardized average
FIGURE 7.14
CLUSTERING BASED ON INTERVAL VARIABLES

GROUP B

B1

GR 110; 12/12 (100%)

Wt: 2.0
Diam: 3.1

B2

GR 120; 6/6 (100%)

Wt: 2.4
Diam: 3.1

GR 110; 18/18 (100%)

B/W: 3.0/2.0 (1.0)
Diam: 4.8

B3

GR 210; 27/27 (100%)

Wt: 1.8
Diam: 4.1

GR 210; 41/36 (88%)

B/W: 2.5/1.8 (0.7)
Diam: 4.7

GR B; 143

B/W: 3.6/2.8 (0.8)
Diam: 8.2

GR 311:

17/N.R.

GR 311:

17/N.R.

GR 310; 44/43 (99%)

B/W: 2.9/2.2 (0.7)
Diam: 4.3

GR 310; 84/71 (89%)

B/W: 2.9/2.4 (0.5)
Diam: 5.1

B4

GR 312; 20/20 (100%)

Wt: 1.9
Diam: 3.2

B5

GR 321; 17/12 (71%)

Wt: 1.7
Diam: 2.6

GR 320; 40/28 (70%)

B/W: 3.3/2.0 (1.3)
Diam: 4.8

B6

GR 322; 17/17 (100%)

Wt: 1.5
Diam: 2.6

ASSESSMENT CLUSTER PROFILES

VARIABLE NUMBERa

<table>
<thead>
<tr>
<th>CLUSTER</th>
<th>1(R)</th>
<th>1(S)b</th>
<th>2(R)</th>
<th>2(S)</th>
<th>3(R)</th>
<th>3(S)</th>
<th>4(R)</th>
<th>4(S)</th>
<th>5(R)</th>
<th>5(S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>1.33</td>
<td>-.23</td>
<td>2.41</td>
<td>.38</td>
<td>2.66</td>
<td>.47</td>
<td>12.16</td>
<td>1.60</td>
<td>1.58</td>
<td>-.83</td>
</tr>
<tr>
<td>B2</td>
<td>2.66</td>
<td>1.33</td>
<td>3.00</td>
<td>.86</td>
<td>3.50</td>
<td>1.31</td>
<td>17.33</td>
<td>2.78</td>
<td>1.83</td>
<td>-.56</td>
</tr>
<tr>
<td>B3</td>
<td>1.03</td>
<td>-.58</td>
<td>.92</td>
<td>-.85</td>
<td>1.59</td>
<td>-.61</td>
<td>1.92</td>
<td>-.73</td>
<td>3.62</td>
<td>1.33</td>
</tr>
<tr>
<td>B4</td>
<td>1.10</td>
<td>-.50</td>
<td>2.85</td>
<td>.74</td>
<td>2.70</td>
<td>.50</td>
<td>3.95</td>
<td>-.27</td>
<td>1.80</td>
<td>-.60</td>
</tr>
<tr>
<td>B5</td>
<td>2.06</td>
<td>.62</td>
<td>1.25</td>
<td>-.58</td>
<td>1.56</td>
<td>-.64</td>
<td>3.31</td>
<td>-.41</td>
<td>2.37</td>
<td>.00</td>
</tr>
<tr>
<td>B6</td>
<td>2.64</td>
<td>1.31</td>
<td>1.62</td>
<td>-.11</td>
<td>2.64</td>
<td>.45</td>
<td>4.70</td>
<td>-.10</td>
<td>1.47</td>
<td>-.95</td>
</tr>
</tbody>
</table>

a As per Figure 5.3
b (R), raw average; (S), standardized average
FIGURE 7.15
OVERLAP ASSESSMENT: 5 INTERVAL VARIABLES

STRESS: 0.05
(4D:0.04; 2D:0.07)
The high within-group cluster overlap indicated a lack of distinctiveness, and suggested that other equally "correct" clusters could have been identified. The lack of between-group cluster replication also indicated that the "correctness" of the clusters was suspect. Thus, no information gathering styles were identified using the five-variable approach.

7.32 4-Variables: Figures 7.16, 7.17, 7.18

In the second interval approach "number of alternatives" was removed from the variable set. As discussed earlier, the "alternatives" questions and the "features" questions were both attempting to measure the type of information gathered. The two categories of questions differed in that the alternatives questions asked "... did you consider...", and the features questions asked for "the main features...considered". Because of this difference, the information gathering as measured by the alternatives questions was potentially less important to the respondent.

Reverse matrix clustering of Group A indicated 19 clusters with high intransiency. The three-cluster level had 87 percent intransient subjects. The five-cluster level had 65 percent, and the ten-cluster level had 89 percent. As was the case in the five-variable Group A approach, there was relatively high intransiency at the small cluster level, high at the general cluster level,
FIGURE 7.16
CLUSTERING BASED ON 4 INTERVAL VARIABLES

GROUP A

A2
GR 2121: 13/13 (100%)
W: 1.4
Diam: 2.6

A3
GR 2122: 21/20 (95%)
W: 1.2
Diam: 2.7

A5
GR 2221: 11/11 (100%)
W: 1.2
Diam: 1.9

A6
GR 2222: 12/10 (.83%)
W: 0.6
Diam: 1.1

A7
GR 311: 14/14 (100%)
W: 1.0
Diam: 1.7

A8
GR 312: 18/18 (100%)
W: 1.4
Diam: 2.4

A9
GR 321: 8/8 (100%)
W: 0.9
Diam: 1.8

A10
GR 322: 11/11 (100%)
W: 1.6
Diam: 2.8

A1
GR 211: 16/9 (56%)
W: 1.2
Diam: 2.0

A4
GR 221: 14/14 (100%)
W: 1.2
Diam: 2.4

A5
GR 222: 23/21 (91%)
W: 1.8/1.0 (0.8)
Diam: 2.8

A6
GR 222: 37/35 (95%)
W: 2.6/1.5 (1.1)
Diam: 3.1

A7
GR 310: 25/14 (40%)
W: 2.6/1.5 (1.1)
Diam: 3.1

A8
GR 320: 19/11 (58%)
W: 2.1/1.3 (0.8)
Diam: 3.5

ASSESSMENT CLUSTER PROFILES

VARIABLE NUMBER

<table>
<thead>
<tr>
<th>CLUSTER</th>
<th>2(R)</th>
<th>2(S)</th>
<th>3(R)</th>
<th>3(S)</th>
<th>4(R)</th>
<th>4(S)</th>
<th>5(R)</th>
<th>5(S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>2.11</td>
<td>.08</td>
<td>1.77</td>
<td>-.63</td>
<td>5.66</td>
<td>.26</td>
<td>2.55</td>
<td>.47</td>
</tr>
<tr>
<td>A2</td>
<td>3.69</td>
<td>1.36</td>
<td>2.53</td>
<td>.22</td>
<td>8.61</td>
<td>.95</td>
<td>2.07</td>
<td>-.02</td>
</tr>
<tr>
<td>A3</td>
<td>3.15</td>
<td>.92</td>
<td>1.95</td>
<td>-.44</td>
<td>2.15</td>
<td>-.55</td>
<td>1.75</td>
<td>-.36</td>
</tr>
<tr>
<td>A4</td>
<td>.93</td>
<td>-.67</td>
<td>1.36</td>
<td>-.12</td>
<td>1.79</td>
<td>-.64</td>
<td>1.50</td>
<td>-.09</td>
</tr>
<tr>
<td>A5</td>
<td>.72</td>
<td>-1.03</td>
<td>1.45</td>
<td>-1.00</td>
<td>2.72</td>
<td>-.42</td>
<td>2.81</td>
<td>.75</td>
</tr>
<tr>
<td>A6</td>
<td>1.30</td>
<td>-.57</td>
<td>2.00</td>
<td>-.38</td>
<td>1.90</td>
<td>-.61</td>
<td>4.00</td>
<td>1.98</td>
</tr>
<tr>
<td>A7</td>
<td>2.50</td>
<td>.39</td>
<td>3.35</td>
<td>1.15</td>
<td>4.07</td>
<td>-.10</td>
<td>1.00</td>
<td>-1.14</td>
</tr>
<tr>
<td>A8</td>
<td>1.66</td>
<td>-.27</td>
<td>2.55</td>
<td>.24</td>
<td>5.83</td>
<td>.30</td>
<td>1.44</td>
<td>-.68</td>
</tr>
<tr>
<td>A9</td>
<td>.62</td>
<td>-1.11</td>
<td>3.12</td>
<td>.88</td>
<td>1.37</td>
<td>-.73</td>
<td>1.75</td>
<td>-.46</td>
</tr>
<tr>
<td>A10</td>
<td>1.00</td>
<td>-.81</td>
<td>3.62</td>
<td>1.46</td>
<td>1.72</td>
<td>-.65</td>
<td>2.81</td>
<td>.75</td>
</tr>
</tbody>
</table>

a As per Figure 5.3
b (R), raw average; (S), standardized average
Although cluster indicates intransience, this link did not occur in BHC

151
FIGURE 7.18
OVERLAP ASSESSMENT: 4 INTERVAL VARIABLES

Stress: 0.066
(4-D:0.00, 2-D:0.151)
but relatively low intransiency at the intermediate cluster level. Consideration of the cluster size statistics for Group A indicated little change in cluster size between the general clusters and the total group cluster. The "average distance" changed from 2.30 to approximately 2.5 standard deviations, and the "diameter" changed from approximately 5.0 to 6.0 standard deviations. This relatively small change in cluster size indicated a lack of cluster distinctiveness at the general cluster level.

Reverse matrix clustering of Group B identified 15 clusters with high intransiency. The three-cluster level had 97 percent intransient subjects. The five-cluster level had 89 percent, and the 11-cluster level had 69 percent. This suggested that the more meaningful clusters were at the intermediate and general levels. This differed from the Group A findings where the meaningful clusters were the small cluster and general cluster levels. As was the case in the Group A analysis, the Group B cluster size statistics indicated a lack of cluster distinctiveness at the general cluster level. For example the average distance within the general clusters was 2.3 standard deviations, and the average distance within the total sample was approximately 2.7.

Nonmetric scaling of the 19 assessment clusters required three dimensions to represent the cluster
locations adequately. Stress in three dimensions was 0.066; in two dimensions, was 0.15. Because of the large number of clusters being considered the TORSCA coordinates for the cluster centroids were plotted by group in Figure 7.18. The dimensions were labeled time, source/visits, and features.

The high within-group overlap and the lack of cluster between-groups replication indicated an absence of meaningful clusters. Thus, no information gathering styles were identified using the four-variable approach.

7.33 3-Variables: Figures 7.19, 7.20, 7.21, 7.22 7.23

In the third approach using interval variables, two variables were removed from the analysis, number of alternatives and number of features. As in the four-variable analysis, the "alternatives" variable was removed because the question wording suggested that it might have been measuring information gathering of relatively low importance to the respondent. The "features" variable was removed because of potential response bias caused by question wording.

Reverse matrix clustering of Group A identified 14 clusters with high intransiency. The three-cluster level had 99 percent intransient subject. The five-cluster level had 96 percent, and the seven assessment clusters had 89 percent. Consideration of cluster size statistics
indicated a relatively large difference in size between the general clusters and the total group. The average distance within the general clusters was 2.1 standard deviations; for the total group the average was approximately 3.5. There was also a relatively large difference between the small clusters and the general clusters - average distances of approximately 1.0 and 2.1 standard deviations respectively. These cluster size differences indicated there were two meaningful levels of clustering, small clusters and general clusters.

Reverse matrix clustering of Group B identified 10 clusters with high intransiency. The three-cluster level had 80 percent intransient subjects. The seven-cluster level had 61 percent. As in the Group A analysis there were relatively large changes in cluster size between the small clusters and the general clusters, and also between the general clusters and the total group. This suggested two meaningful cluster levels.

Overlap assessment of 15 clusters from Groups A and B required two dimensions to describe the cluster locations. Stress in two dimensions was 0.056. The two dimensions were labeled visits/sources, and time, (Figure 7.22). The unscaled centroid coordinates were also plotted to check for possible differences between two and three dimensional solutions (Figure 7.23). Since there appeared to be little difference in relative cluster
FIGUR: 7.19
CLUSTERING BASED ON 3 INTERVAL VARIABLES
GROUP A

A2
GR 221: 20/19 (95%)
W: 0.9
Diam: 1.4

A1
GR 210: 10/16 (89%)
W: 1.7
Diam: 3.0

A3
GR 222: 11/10 (91%)
W: 0.8
Diam: 1.4

GR 220: 31/31 (100%)
B/W: 1.4/1.9 (0.5)
Diam: 2.1

GR 200: 49/47 (96%)
B/W: 2.2/1.3 (0.9)
Diam: 3.9

A4
GR 321: 39/36 (92%)
W: 1.1
Diam: 2.3

GR 310: 7/7 (100%)
W: 1.6
Diam: 2.7

GR 300: 92/92 (100%)
B/W: 2.0/1.6 (1.2)
Diam: 4.6

GR 144
B/W: 5.0/2.1 (2.9)
Diam: 7.4

A5
GR 322: 18/18 (100%)
W: 0.9
Diam: 1.8

GR 322: 14/12 (86%)
W: 0.6
Diam: 1.1

A6
GR 323: 14/12 (86%)
W: 0.3
Diam: 0.7

ASSESSMENT CLUSTER PROFILES

<table>
<thead>
<tr>
<th>VARIABLE NUMBER</th>
<th>CLUSTER 2(R)b</th>
<th>CLUSTER 2(S)</th>
<th>CLUSTER 4(R)</th>
<th>CLUSTER 4(S)</th>
<th>CLUSTER 5(R)</th>
<th>CLUSTER 5(S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>3.43</td>
<td>1.15</td>
<td>8.81</td>
<td>1.00</td>
<td>2.75</td>
<td>.68</td>
</tr>
<tr>
<td>A2</td>
<td>3.30</td>
<td>1.04</td>
<td>2.35</td>
<td>-.50</td>
<td>1.60</td>
<td>-.51</td>
</tr>
<tr>
<td>A3</td>
<td>3.50</td>
<td>1.22</td>
<td>6.60</td>
<td>.49</td>
<td>1.50</td>
<td>-.62</td>
</tr>
<tr>
<td>A4</td>
<td>1.41</td>
<td>-.47</td>
<td>3.13</td>
<td>-.32</td>
<td>1.41</td>
<td>-.70</td>
</tr>
<tr>
<td>A5</td>
<td>1.06</td>
<td>-.76</td>
<td>2.28</td>
<td>-.52</td>
<td>3.40</td>
<td>1.36</td>
</tr>
<tr>
<td>A6</td>
<td>.91</td>
<td>-.88</td>
<td>1.33</td>
<td>-.74</td>
<td>2.00</td>
<td>-.10</td>
</tr>
<tr>
<td>A7</td>
<td>3.33</td>
<td>1.07</td>
<td>23.33</td>
<td>4.39</td>
<td>1.00</td>
<td>-1.14</td>
</tr>
</tbody>
</table>

a As per Figure 5.3
b (R), raw average; (S), standardized average
FIGURE 7.21
OVERLAP ASSESSMENT: 3 INTERVAL VARIABLES
DISTINCTIVENESS
FIGURE 7.22
OVERLAP ASSESSMENT: 3 INTERVAL VARIABLES
REPLICATION

Stress: 0.056
(3-D:0.00, 1-D:0.188)

<table>
<thead>
<tr>
<th>THOROUGH</th>
<th>INFORMATION SOURCES</th>
<th>STORE VISITS</th>
<th>TIME SPENT</th>
<th>GROUP SIZE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GR A</td>
<td>GR B</td>
<td>GR A</td>
<td>GR B</td>
</tr>
<tr>
<td>THOROUGH</td>
<td>(A7-B1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MODERATELY THOROUGH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEEKS</td>
<td>(A1-B6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MONTHS</td>
<td>(A3-B7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MONTHS PLUS</td>
<td>(A2-B5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NON-THOROUGH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAYS PLUS</td>
<td>(A5-B2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEEK-MONTHS</td>
<td>(A6-B3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MONTHS PLUS</td>
<td>(A4-B4)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 7.23
OVERLAP ASSESSMENT: 3 INTERNAL VARIABLES
3-DIMENSIONAL
locations between the two solutions the discussion centers on the two dimensional plot.

Three main groupings of clusters were identified. They corresponded to the three general clusters identified in reverse matrix analyses. One grouping included clusters A7 and B1; the second included clusters A1, A2, A3 and B5, B6, B7; the third included clusters A4, A5, A6 and B2, B3, B4. These groupings had low within-group overlap, and high between-group overlap, indicating good cluster distinctiveness and good cluster replication. The cluster profiles in the three groupings suggested prepurchase styles that differed with respect to number of store visits and to a lesser extent number of information sources. The cluster profiles arranged to facilitate comparison across the three groupings is presented in Figure 7.22. The "thorough" style (A7-B1) represented slightly more than three information sources, approximately 20 store visits, and approximately a year to come to the decision. In Groups A and B combined there were only 11 subjects in this style category. The "moderately thorough" style represented slightly more than three information sources, three to ten store visits, and varied considerably with respect to "time spent". There were 99 subjects in this style category. Finally, the "non-thorough" style represented approximately one information source, one to three store visits, and varied considerably
with respect to "time spent". There were 114 subjects in this style category.

Within the two large cluster groupings, moderately thorough and non-thorough, small clusters formed substyles that differed in terms of time. In the moderately thorough style 26 subjects (clusters A1 and B6) spent "weeks" gathering information, 31 subjects (clusters A3 and B7) spent "months", and 42 subjects (clusters A2 and B5) spent "months plus". In the non-thorough style 43 subjects (clusters A5 and B2) spent "days plus", 35 subjects spent "weeks to months", and 46 subjects spent "months plus". A graphical representation of the six substyles is presented below.

<table>
<thead>
<tr>
<th>STYLE</th>
<th>MODERATELY THOROUGH</th>
<th>TIME SPENT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DAYS</td>
<td>WEEKS</td>
</tr>
<tr>
<td>&quot;fast&quot; (A1-B7)</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>&quot;moderate&quot; (A3-B8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;slow&quot; (A2-B5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NON-THOROUGH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;fast&quot; (A5-B2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;moderate&quot; (A6-B3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;slow&quot; (A4-B4)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7.34 Differences in Corresponding Clusters Between Group A and Group B

The final consideration with respect to the three-variable analysis was to attempt to account for the
differences between Group A and Group B in the number of respondents in corresponding clusters. Since the two subsamples, Group A and Group B, were randomly taken from the total furniture sample, little difference in number of subjects in corresponding clusters was expected.

The differences noted were: first, there was a difference in the percent of intransient subjects in the general clusters. In Group A there were 99 percent intransient subjects at the three-cluster level; in Group B there were 80 percent. Second, there were differences between groups in the number of subjects in the three major styles. There were 80 subjects from Group A and 34 from Group B in the "non-thorough" style; there were 45 and 54 from Groups A and B in the "moderately thorough" style. Third, there were differences between groups in the number of subjects in the substyles. For example, there were 32 subjects from Group A and 11 subjects from Group B in the "non-thorough/fast" substyle.

The differences between groups probably resulted from the presence of "border line" subjects. A border line subject was interpreted conceptually as indicating that a subject (or cluster) was equi-distant between
two or more other subjects (or clusters). The result of this condition was that the subject could be clustered "correctly" with more than one cluster. Border line subjects were defined as transients. Since the identification of information gathering styles was based on intransient clusters, it was useful to consider two characteristics of the identification of intransients that could have resulted in differences in cluster sizes between groups.

First, for a subject to be identified as transient his distance from two or more other subjects had to be exactly the same. Consider the two subject clusters plotted below. In case A the center group of subjects is exactly equi-distant between the two main clusters. For a two cluster solution the center group is identified as transients. Thus, the intransient clusters each contain 39 percent of the sample - a total of 78 percent intransients.

**CASE A**

```
       X

Subject Scatter
```

```
Cluster 1- 39% of sample

Cluster 2- 39%

Intransient Clusters

2-Cluster Solution
```
In case B the center group of subjects is slightly closer to the right hand cluster. For a two cluster solution the center group cluster with the right hand cluster, so that the intransient clusters would contain 39 and 61 percent of the sample - a total of 100 percent intransients. Thus, the first potential cause of between-group differences in number of subjects in corresponding clusters was - slight differences in intersubject distances were sufficient to influence the cluster membership of border line subjects.

The second characteristic of intransient subject identification potentially resulting in differences was that all transients were not necessarily identified. Cluster membership of the border line subjects was determined by the subject-order in matrix. Reverse matrix clustering was used to help identify border line subjects by clustering the proximity matrix using two different subject orders. Tests with known data indicated successful identification of transient subjects using the reverse matrix procedure. However, as
suggested earlier with complex data (for example multi-way ties) multi-order matrix clustering would be required to identify all transient subjects. Thus, differences in the extent to which actual intransients were identified provided a second potential cause of differences between groups. For example, in one group the transients in a particular area of the subject scatter may be identified well and removed from the final clusters. For the other group the transients in the same area of the scatter may be identified poorly and thus, be included in the final clusters.

7.35 Interval Variable Approach: Summary

Three approaches were based on the five interval variables. The first included all five variables. In the second the "alternatives" variable was removed since the wording of the questions allowed endorsement even when the alternatives were relatively unimportant to the respondent. The third iteration removed both the "alternatives" and "features" variables, the latter because of potential response bias.

Neither the five-variable nor the four-variable approach led to the identification of information gathering styles. High within-group overlap and lack of between-group replication suggested a relatively uniform distribution of subjects, and thus, a lack of
meaningful clusters.

The three-variable approach identified replicating clusters at two levels of generality. General clusters indicated three major information gathering styles, thorough, moderately thorough, and non-thorough. The three styles differed primarily in terms of number of store visits, and number of information sources. Although the thorough style represented only four percent of the sample, the moderately thorough and non-thorough styles represented substantial groupings, 34 and 40 percent of the sample respectively. Within each of the two major styles three substyles were identified. The three differed in terms of "time". For example within the "non-deliberate" style there were substyles that spent "days", "week to months", and "months plus" on the decision making process. A diagram of these styles is presented below.

```
FURNITURE BUYERS
  /
 /  
Thorough  Moderately Thorough  Non-Thorough
  /     /    
Months Plus  Months  Weeks  Months Plus
       /    /     
       Months Weeks MONTHS  Days
```
7.4 RE-TEST ANALYSIS

The purpose of the exploration analysis was to study the prepurchase information gathering of furniture buyers to develop a model of information gathering, and to identify the "best" approach(s) to modeling. Of the seven approaches tested two provided the "best" results. The 14-variable nominal approach identified two dimensions of information gathering, stores and time, and identified two general styles differing primarily in terms of time. The three-variable interval approach also identified stores and time as two major dimensions. In this case the styles identified had two levels of generality. There were three general styles, thorough, moderately thorough, and non-thorough. Each of the latter two contained three substyles, short, moderate and long time.

Comparison of the furniture and appliance sample means, Table 7.7, indicated that appliance buyers were generally less thorough. Thus, the clustering results were not expected to be identical for the two samples. The appliance sample was expected to have more non-thorough clusters or more subjects in the non-thorough clusters. Since the non-thorough nature of appliance buyers was relatively general across variables, the approaches that had proven "best" for clustering furniture buyers were expected to be best for
<table>
<thead>
<tr>
<th>Alternatives</th>
<th>Furniture</th>
<th>Appliances</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Save/Spend</td>
<td>0.31</td>
<td>0.25a</td>
</tr>
<tr>
<td>2 Brand (wide open)</td>
<td>0.79</td>
<td>0.57a</td>
</tr>
<tr>
<td>3 Price</td>
<td>0.40</td>
<td>0.34</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sources</th>
<th>Furniture</th>
<th>Appliances</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Friends</td>
<td>0.17</td>
<td>0.35a</td>
</tr>
<tr>
<td>5 Salesmen</td>
<td>0.38</td>
<td>0.39</td>
</tr>
<tr>
<td>6 Stores</td>
<td>0.57</td>
<td>0.38a</td>
</tr>
<tr>
<td>7 Advertisements</td>
<td>0.31</td>
<td>0.30</td>
</tr>
<tr>
<td>8 Other</td>
<td>0.17</td>
<td>0.15</td>
</tr>
<tr>
<td>9 Family</td>
<td>0.36</td>
<td>0.31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Features</th>
<th>Furniture</th>
<th>Appliances</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 Price</td>
<td>0.57</td>
<td>0.56a</td>
</tr>
<tr>
<td>11 Brand</td>
<td>0.02</td>
<td>0.12a</td>
</tr>
<tr>
<td>12 Style</td>
<td>0.71</td>
<td>0.30a</td>
</tr>
<tr>
<td>13 Quality</td>
<td>0.48</td>
<td>0.16a</td>
</tr>
<tr>
<td>14 Size</td>
<td>0.47</td>
<td>0.58a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Store Visits</th>
<th>Furniture</th>
<th>Appliances</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 Total stores</td>
<td>0.60</td>
<td>0.31a</td>
</tr>
<tr>
<td>16 Maximum visits to single store</td>
<td>0.42</td>
<td>0.28a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>Furniture</th>
<th>Appliances</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 Time</td>
<td>2.23</td>
<td>2.60a</td>
</tr>
</tbody>
</table>

1 Number of Alternatives | 1.51 | 1.17a |
2 Number of Sources      | 1.97 | 1.85  |
3 Number of Features     | 2.27 | 1.75a |
4 Total Visits           | 4.83 | 2.67a |

*Significant Difference between samples at α of 0.05.*
appliance buyers. Thus, the 14-variable nominal and three-variable interval approaches were applied to the appliance data.

There were two major exceptions to the less thorough nature of appliance buyers. Only 57 percent of appliance buyers considered alternative brands to be "wide open" - 79 percent of furniture buyers. Second, 35 percent of appliance buyers obtained information from friends - 17 percent of furniture buyers. Because alternative brands and information from friends were of greater significance for appliance buyers, these two variables were given special consideration. After analysing the appliance data using the 14-variable and three-variable approaches, analysis was done to test the effect of adding "alternative brands" or "information from friends" to the three-variable interval approach.

7.41 14-Variables Re-Test

The 14-nominal analysis of the appliance data did not lead to the identification of search styles; thus, it is not reported in detail. Analyses were done on both halves of a random split of the appliance data. In reverse matrix clustering of Group B all clusters, both large and small had low intransiency. In every case but one less than one half of the forward-matrix cluster members remained in a common cluster in the reverse
matrix analysis. This suggested a uniform distribution of subjects, and thus, a lack of meaningful clusters.

7.42 3-Variables Re-Test: Figures 7.24, 7.25, 7.26, 7.27

Product moment correlation of the three variables, Table 7.8, indicated a relatively uniform degree of correlation between variable-pairs. Thus, each variable would have equal impact on the inter-subject distance matrix. The maximum common variance between variables was 22 percent indicating that redundant variables were not a problem.

Subgroup comparisons for the appliance data are presented in Table 7.9. Comparison based on purchase date indicated no consistent significant differences, although there was a consistent tendency for recent buyers to report more deliberate search behavior. There was one consistent difference between "new" and "used" categories with the average new purchase being considered for a longer time (2.51 vs. 3.23). The one other consistent difference was between washing machines and other products with the average washing machine purchase being considered for a shorter time. These differences were judged to be relatively minor, and were not expected to result in subgroups with distinctive search styles.

Reverse matrix clustering of Group A identified 10
TABLE 7.8
PRODUCT MOMENT CORRELATION:
3 INTERVAL VARIABLES:
APPLIANCES DATA

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 Information Sources</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Visits</td>
<td>.47&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.31</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>5 Time&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-.43</td>
<td>-.35&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>-.27</td>
<td>-.29</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<sup>a</sup> N of 130, r of ±.17 significantly different from 0 at α of 0.05

<sup>b</sup> Time is coded 1, 2, 3, 4 for years, months, weeks and days

<sup>c</sup> Upper correlation: Group A, lower correlation: Group B
TABLE 7.9
SUB-GROUP COMPARISONS:
3 INTERVAL VARIABLES:
APPLIANCE DATA

<table>
<thead>
<tr>
<th>PURCHASE DATEc</th>
<th>Information Sources</th>
<th>Store Visits</th>
<th>Timed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>Pre July'69</td>
<td>1.79</td>
<td>1.72</td>
<td>2.52</td>
</tr>
<tr>
<td>(98) (91)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post June'69</td>
<td>2.28</td>
<td>1.79</td>
<td>3.11</td>
</tr>
<tr>
<td>(42) (49)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEW-USED</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>USED</td>
<td>1.10</td>
<td>1.36</td>
<td>1.89</td>
</tr>
<tr>
<td>(19) (19)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEW</td>
<td>2.07</td>
<td>1.80</td>
<td>2.82</td>
</tr>
<tr>
<td>(121) (121)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRODUCT</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>Washing Machine</td>
<td>1.79</td>
<td>2.00</td>
<td>2.29</td>
</tr>
<tr>
<td>(34) (34)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clothes Dryer</td>
<td>2.25</td>
<td>1.85</td>
<td>1.87</td>
</tr>
<tr>
<td>(7) (8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stove</td>
<td>1.82</td>
<td>1.26</td>
<td>2.64</td>
</tr>
<tr>
<td>(15) (17)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refrigerator</td>
<td>1.64</td>
<td>1.60</td>
<td>4.00a</td>
</tr>
<tr>
<td>(25) (17)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Food Freezer</td>
<td>2.55</td>
<td>2.81a</td>
<td>3.44</td>
</tr>
<tr>
<td>(11) (9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colour Television</td>
<td>1.94</td>
<td>1.85</td>
<td>2.35</td>
</tr>
<tr>
<td>(14) (17)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Air Conditioner</td>
<td>1.75</td>
<td>1.66</td>
<td>1.87</td>
</tr>
<tr>
<td>(3) (8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black and White TV</td>
<td>1.69</td>
<td>1.32</td>
<td>2.30</td>
</tr>
<tr>
<td>(25) (26)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dish Washer</td>
<td>5.00a</td>
<td>1.66</td>
<td>6.50a</td>
</tr>
<tr>
<td>(6) (4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL GROUP STATISTICS</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>Averages</td>
<td>1.94</td>
<td>1.75</td>
<td>2.70</td>
</tr>
<tr>
<td>Standard Deviations</td>
<td>1.36</td>
<td>1.24</td>
<td>2.30</td>
</tr>
</tbody>
</table>

a Significant at $\alpha$ of 0.05
b Significant at $\alpha$ of 0.10
c Survey was conducted in November 1969
d Time is coded 1,2,3,4 for years, months, weeks and days
clusters with high intransiency. The three-cluster level had 69 percent intransient subjects, and the six assessment clusters had 96 percent. The increase in percent intransients between the general clusters, and the small clusters was due to cluster A2. Although cluster A2 had high intransience, it did not cluster consistently at the general cluster level. In the forward matrix analysis A2 clustered with A5 at the general level. In the reverse matrix analysis it clustered with A3. (Figure 7.26 shows considerable overlap between A2 and A5, and between A2 and A3.) The cluster statistics indicated a relatively large change in size between the general clusters and the total sample, suggesting meaningful differences among the three general clusters.

Reverse matrix clustering of Group B identified nine clusters with high intransiency. The three-cluster level had 73 percent intransient subjects, and the six assessment clusters had 83 percent. Once again one cluster resulted in the difference in percent intransients. In forward matrix clustering B2 clustered with B6 at the general cluster level; in reverse matrix clustering B2 clustered with B1. (Figure 7.26 shows considerable overlap between B2 and B6, and between B2 and B1.) Cluster statistics indicate a relatively large increase in cluster size between the general cluster level and the total sample, suggesting meaningful differences among
FIGURE 7.34
CLUSTERING BASED ON 3 INTERVAL VARIABLES

GROUP A: RE-TEST

A5
GR 221: 14/14 (100%)
W: 0.7
Diam: 1.3

A2
GR 222: 41/36 (88%)
W: 1.1
Diam: 3.0

A6
GR 230: 33/33 (100%)
W: 0.8
Diam: 2.0

GR 310: 13/13 (100%)
W: 0.8
Diam: 2.8

GR 320: 15/15 (100%)
W: 1.5
Diam: 2.8

A1
GR 10C: 7/7 (100%)
W: 1.8
Diam: 3.3

A3
GR 300: 28/28 (100%)
W: 1.8
Diam: 3.1

GR A: 132
B/W: 2.7/1.7 (1.0)
Diam: 6.3

ASSESSMENT CLUSTER PROFILES

<table>
<thead>
<tr>
<th>VARIABLE NUMBER</th>
<th>CLUSTER 2(R)b</th>
<th>2(S)</th>
<th>4(R)</th>
<th>4(S)</th>
<th>5(R)</th>
<th>5(S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>1.85</td>
<td>-1.12</td>
<td>8.57</td>
<td>2.48</td>
<td>1.71</td>
<td>-0.90</td>
</tr>
<tr>
<td>A2</td>
<td>1.88</td>
<td>-1.10</td>
<td>1.41</td>
<td>-0.56</td>
<td>1.86</td>
<td>-0.76</td>
</tr>
<tr>
<td>A3</td>
<td>3.96</td>
<td>1.44</td>
<td>4.57</td>
<td>.78</td>
<td>1.85</td>
<td>-0.76</td>
</tr>
<tr>
<td>A4</td>
<td>2.33</td>
<td>.22</td>
<td>5.33</td>
<td>1.10</td>
<td>3.22</td>
<td>.61</td>
</tr>
<tr>
<td>A5</td>
<td>.64</td>
<td>-1.03</td>
<td>1.07</td>
<td>-0.70</td>
<td>2.71</td>
<td>.10</td>
</tr>
<tr>
<td>A6</td>
<td>1.12</td>
<td>-0.67</td>
<td>1.39</td>
<td>-0.57</td>
<td>4.00</td>
<td>1.40</td>
</tr>
</tbody>
</table>

a As per Figure 5.3
b (R), raw average; (S), standardized average
c Although this cluster indicates i.transience, this link did not occur in reversed clustering.
Figure 7.25
Clustering Based on 3 Interval Variables
Group B: IE-Test

B2
GR 210: 16/13 (81%)
Wt: 1.6
Diam: 3.5

B1
GR 100: 11/11 (100%)
Wt: 1.8
Diam: 3.7

C

B6
GR 220: 15/14 (93%)
Wt: 1.3
Diam: 2.5

GR 200: 31/14 (45%)
B/W: 2.1/1.5 (0.7)
Diam: 4.2

GR B: 127
B/W: 2.7/1.6 (1.1)
Diam: 5.9

GR 310:
B/W: 8/(N.R.)

GR 300: 85/68 (80%)
B/W: 2.1/1.2 (0.8)
Diam: 3.5

B3
GR 331: 23/18 (78%)
Wt: 1.2
Diam: 2.3

B4
GR 320: 27/23 (85%)
Wt: 0.9
Diam: 1.6

B5
GR 332: 27/27 (100%)
Wt: 1.2
Diam: 2.0

Assessment Cluster Profiles

<table>
<thead>
<tr>
<th>Variable Number</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>B4</th>
<th>B5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster</td>
<td>2(R)</td>
<td>2(S)</td>
<td>4(R)</td>
<td>4(S)</td>
<td>5(R)</td>
</tr>
<tr>
<td>2.63</td>
<td>.67</td>
<td>7.45</td>
<td>2.20</td>
<td>1.90</td>
<td>- .64</td>
</tr>
<tr>
<td>3.53</td>
<td>1.33</td>
<td>4.69</td>
<td>.94</td>
<td>1.30</td>
<td>-1.22</td>
</tr>
<tr>
<td>1.22</td>
<td>-.37</td>
<td>1.11</td>
<td>-.68</td>
<td>1.61</td>
<td>-.93</td>
</tr>
<tr>
<td>.73</td>
<td>-.73</td>
<td>1.13</td>
<td>-.67</td>
<td>4.00</td>
<td>1.36</td>
</tr>
<tr>
<td>.88</td>
<td>-.62</td>
<td>1.11</td>
<td>-.68</td>
<td>2.88</td>
<td>.29</td>
</tr>
<tr>
<td>3.42</td>
<td>1.25</td>
<td>2.21</td>
<td>-.18</td>
<td>2.85</td>
<td>.29</td>
</tr>
</tbody>
</table>

a As per Figure 5.3
b (R), raw average; (S), standardized average
c Although this cluster indicates intransience, this link did not occur in reverse clustering.
<table>
<thead>
<tr>
<th>SOURCES</th>
<th>VISITS</th>
<th>TIME</th>
<th>GROUP SIZE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>THOROUGH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A1-B1)</td>
<td>1.85</td>
<td>2.63</td>
<td>8.57 7.45</td>
</tr>
<tr>
<td></td>
<td>1.71</td>
<td>1.90</td>
<td>7 11</td>
</tr>
<tr>
<td>MODERATE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A3-B2)</td>
<td>3.96</td>
<td>3.54</td>
<td>4.57 4.69</td>
</tr>
<tr>
<td></td>
<td>1.85</td>
<td>1.30</td>
<td>28 13</td>
</tr>
<tr>
<td>NON-THOROUGH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A6-B4)</td>
<td>1.12</td>
<td>.73</td>
<td>1.39 1.13</td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>4.00</td>
<td>33 23</td>
</tr>
<tr>
<td>(A5-B5)</td>
<td>.64</td>
<td>.88</td>
<td>1.07 1.11</td>
</tr>
<tr>
<td></td>
<td>2.71</td>
<td>2.88</td>
<td>14 27</td>
</tr>
<tr>
<td>(A2-B3)</td>
<td>1.88</td>
<td>1.22</td>
<td>1.41 1.11</td>
</tr>
<tr>
<td></td>
<td>1.86</td>
<td>1.61</td>
<td>36 18</td>
</tr>
</tbody>
</table>
the three general clusters.

The stress in two dimensions for the 12 assessment clusters was "fair" (0.09). Thus, reduced-space analysis was not used. Figures 7.26 and 7.27 present cluster centroid locations based on the three standardized interval variables. Although Figure 7.26 indicated high within-group overlap, the between-group replication in Figure 7.27 indicated that there were meaningful concentrations of subjects in the subject space. The 12 assessment clusters formed 5 replicating pairs. The replicating clusters suggested three styles differing in terms of "store visits", and to a lesser extent "number of sources". A1-B1 represented many visits and a moderate number of sources; B2-A3 represented moderate visits and many sources; A2-B3, A5-B5, A6-B4 represented few visits and few sources. The latter three cluster pairs also differed in terms of time; thus they represented 3 substyles, "months plus", "weeks plus", and "days".

The two non-replicating clusters had relatively few cluster members. With a larger sample they might have been closer together, or complimentary clusters might have been identified in each subsample.

In summary the re-test analysis evaluated the appliance data using two approaches, 14-variable nominal and three-variable interval. The 14-variable nominal approach did not prove successful, since in one half of
the sample no clusters with high intransiency were identified. The three-variable interval approach led to the identification of information gathering styles at two levels of generality. Three general styles differed in terms of "store visits", and "number of sources". Two were relatively thorough, however, differed in terms of emphasis on store visits. The "non-thorough" style, few visits/few sources, contained three substyles differing in terms of "time spent". A diagram of the three-variable interval styles is presented below.

```
    APPLIANCE BUYERS
        /              \
       /                \
    Thorough:         Thorough:       Non-Thorough
        Balanced       Store Intense
                         /           /
                    Months Plus  Weeks-Months  Days
```

7.43 Consideration of Differences Between Furniture and Appliance Buyers

For the most part appliance buyers were less thorough than furniture buyers. However, there were two sample means that were clearly different from the trend. In the consideration of alternative brands, only 57 percent of appliance buyers reported the choice to be "wide open" - 79 percent of furniture buyers. Second, 35 percent of appliance buyers obtained information from friends - 17
percent of furniture buyers. It was considered important to see if including either of these two variables would lead to the identification of information gathering styles.

The effect of including "alternative brands" was tested by clustering appliance buyers based on the three interval variables (sources, visits and time) plus alternative brands. When alternative brands was used in the furniture data analyses, it was used in a dichotomous format, one/few brands vs. wide open. This was done because of the low endorsement of the "one" and "few" categories (8 and 13 percent of the sample). For appliance buyers the distribution of responses across one, few and many brands was 24, 19 and 57 percent. Therefore, when alternative brands was added to the three variable approach, it was added as an interval variable with three possible responses.

The "three-variable plus alternative brands" approach did not lead to the identification of information gathering styles that replicated across Group A and Group B. The distribution of clusters suggested a relatively uniform distribution of subjects in the subject-space. To further assess the influence of alternative brands, the response distribution to this variable was determined for each of the three-variable approach styles. As indicated below the response
distribution in each style matched closely the distribution for the total appliance sample.

### 259 Appliance Buyers

<table>
<thead>
<tr>
<th>Brand</th>
<th>% of Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>open</td>
<td>55.6</td>
</tr>
<tr>
<td>few</td>
<td>20.0</td>
</tr>
<tr>
<td>one</td>
<td>24.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Many Visits</th>
<th>Many Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand</td>
<td>% of Sample</td>
</tr>
<tr>
<td>open</td>
<td>55.6</td>
</tr>
<tr>
<td>few</td>
<td>33.3</td>
</tr>
<tr>
<td>one</td>
<td>11.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Moderate Visits</th>
<th>Many Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand</td>
<td>% of Sample</td>
</tr>
<tr>
<td>open</td>
<td>68.2</td>
</tr>
<tr>
<td>few</td>
<td>22.0</td>
</tr>
<tr>
<td>one</td>
<td>9.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Few Visits/Few Sources</th>
<th>Slow</th>
<th>Moderate</th>
<th>Fast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand</td>
<td>% of Sample</td>
<td>% of Sample</td>
<td>% of Sample</td>
</tr>
<tr>
<td>open</td>
<td>54.7</td>
<td>56.1</td>
<td>62.5</td>
</tr>
<tr>
<td>few</td>
<td>13.2</td>
<td>17.1</td>
<td>17.9</td>
</tr>
<tr>
<td>one</td>
<td>32.1</td>
<td>26.8</td>
<td>19.6</td>
</tr>
</tbody>
</table>

The relatively uniform distribution of alternative brands across styles helped to explain why distinctive clusters were not identified in the "three-variable plus alternative brands" approach, although they were in the "three-variable" approach. Since the distribution of
alternative brand in each three-variable style was approximately the same, the addition of the brand variable in effect reduced the differences between styles. In other words including brand resulted in a more uniform distribution in the subject scatter.

The distribution of "information from friends" was determined for each three-variable style. The proportion of each style that considered information from friends matched closely the proportion of the total sample. Thus, addition of information from friends was not expected to lead to the identification of information gathering styles.

One reason for considering alternative brands and information from friends was to see if they would lead to identification of information gathering styles. Second, considering these variables served to point out the importance of comparing cluster members on variables not included in the proximity calculations. The styles groups identified were similar in terms of the three interval variables (sources, visits, and time). However, in order to understand or make use of the clusters, comparison of the subject members on other variables is required. For example in the "few visits/short time" style of the three-variable appliance analysis 20 percent considered only one brand and 62 percent considered brand "wide open". This suggests
that although these subjects used the same visits/time style, two different brand strategies would be required to attract them to a product. Comparisons on other variables are discussed further in Chapter 8.
CHAPTER 8
SUMMARY AND IMPLICATIONS

8.1 RESEARCH SUMMARY

The basic purpose of this research was to develop a descriptive model of the prepurchase information gathering process of household durable buyers. The two central questions were: was it possible to identify major dimensions of information gathering? Second, was it possible to identify major information gathering styles? Three steps were involved in the modeling process. The first was assessment and selection of variables describing information gathering. The second was assessment and development of numerical taxonomic procedures. Finally, numerical taxonomic analysis of the information gathering variables led to identification of dimensions and styles useful in describing the information gathering process. This chapter contains four major sections. The first section summarizes the research analysis and findings. The second, third and fourth sections discuss the three major areas of study – modeling prepurchase information gathering, measurement of prepurchase information gathering, and numerical taxonomic procedures. For
each area of study the discussion centers on conclusions with respect to the findings, and implications for future research.

8.11 Assessment and Selection of Variables

The Fry-Portis survey of household durable buyers was conducted to increase understanding of buyer information gathering. The survey included 25 variables describing information gathering. Variables were removed from the set of 25 because of lack of discrimination across subjects, or excessive redundancy among the measures, or potential response bias resulting from question wording. The preliminary analysis resulted in a final set of 18 variables. Three variables dealt with alternatives considered, six with information sources used, five with features considered, three with store visits, and one with time spent. Sixteen of the variables were dichotomous; one had three categories; one had four categories.

In keeping with the exploratory nature of the research these variables were tested using two approaches. For one approach subject comparisons were based on the 18 variables. There was no preliminary weighting or aggregation of the variables. This was termed the nominal approach. A simplified form of analysis was provided by aggregating the 18 variables to form five interval
variables. This was termed the interval approach, and included the following variables: number of alternatives, number of sources, number of features, number of visits, and time. Both approaches were tested since it was not apparent which would be more fruitful in modeling pre-purchase information gathering.

8.12 Numerical Taxonomic Procedures

Preliminary analysis of numerical taxonomic procedures considered three factors - selection of proximity measures, selection of a clustering algorithm, and selection and development of procedures for evaluating clustering results. The proximity measures selected were determined by the type of variables used. City-block distance was selected for use with the 18 nominal variables. It was selected since, relative to similarity, it increased the potential impact of the time variable. Euclidean distance was selected for use with the 5 interval variables. It was selected since it did not remove the influence of level and scatter, whereas correlation did.

The clustering algorithm selected was Johnson's "diameter" method (Johnson, 1967). This algorithm was chosen for three reasons. It identified compact clusters; it required only rank order assumptions of the proximity matrix; and the clustering output was hierarchical.
Due to a lack of definitive procedures for evaluating clustering results, the assessment of numerical taxonomic procedures centered in this area. The first problem addressed was the effect of proximity ties\(^1\) on cluster membership when using the diameter algorithm. It was demonstrated that subjects associated with the proximity ties could be clustered "correctly" in more than one cluster. As part of this research a procedure was developed to reduce the possibility of ambiguous results. The process was termed reverse matrix clustering. After the initial clustering the subject-order in the proximity matrix was reversed, and the subjects reclustered. Inconsistencies in cluster membership between the two clusterings were identified and were termed transients.

A second problem associated with numerical taxonomic procedures was identification of the most meaningful number of clusters. Two methods of selecting the "best" number were examined. The first method involved a plot comparing cluster diameter and number of clusters. The second method was a comparison of average distance within clusters and average distance between clusters. A weakness in these methods was a lack of clear indication of

---

\(^1\) Equal inter-subject distances in the distance matrix was referred to as a proximity tie.
cluster distinctiveness; that is, they provided little indication of the relative similarity or dissimilarity of the "best" clusters.

The second procedure developed during the research was overlap assessment. It was aimed at providing a guide to cluster distinctiveness and replication. This procedure involved four steps. Given a set of clusters of interest, the first step was to calculate a distance matrix of distances between each pair of cluster centroids. Cluster centroid was defined as the average, across cluster members, on each variable used for inter-subject comparison. The second step was to analyse the centroid distance matrix using nonmetric multidimensional scaling. Third, the centroid configuration was plotted in the appropriate dimensionality. Fourth, an indication of relative cluster size was obtained by plotting the appropriate cluster diameter at each scaled centroid location. Relative cluster distinctiveness was assessed by considering the overlap of neighbouring clusters. Since it was possible to include clusters from both halves of split-sample analyses, overlap assessment also provided an indication of cluster replication.

8.13 Numerical Taxonomic Analysis of Prepurchase Information Gathering

Numerical taxonomic analysis of prepurchase information gathering was done in two segments. The exploration
segment was based on data describing furniture buyers. This segment tested a variety of approaches that differed in terms of number and form of variables used. In the re-test segment the approaches that had been most fruitful in the exploration segment were repeated using appliance buyer data. A list of approaches and findings is presented below.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exploration Segment</td>
<td></td>
</tr>
<tr>
<td>18 Nominal Variables</td>
<td>2 general styles differing in terms of store visits</td>
</tr>
<tr>
<td>14 Nominal Variables</td>
<td>2 general styles differing in terms of store visits and time</td>
</tr>
<tr>
<td>12 Nominal Variables</td>
<td>No styles identified</td>
</tr>
<tr>
<td>9 Nominal Variables</td>
<td>No styles identified</td>
</tr>
<tr>
<td>5 Interval Variables</td>
<td>No styles identified</td>
</tr>
<tr>
<td>4 Interval Variables</td>
<td>No styles identified</td>
</tr>
<tr>
<td>3 Interval Variables</td>
<td>7 styles differing in terms of visits, sources and time</td>
</tr>
<tr>
<td>Re-test Segment</td>
<td></td>
</tr>
<tr>
<td>14 Nominal Variables</td>
<td>No styles identified</td>
</tr>
<tr>
<td>3 Interval Variables</td>
<td>5 styles differing in terms of visits, sources and time</td>
</tr>
</tbody>
</table>

Two approaches in the exploration segment were of primary interest in terms of information gathering style identification. First, in the 14-variable nominal
approach (with "alternatives" and "total visits" removed) two general information gathering styles were identified - deliberate and non-deliberate. Second, in the three-variable interval approach ("alternatives" and "features" removed) several information gathering styles were identified. There were three general styles that differed primarily with respect to number of store visits and number of sources. In two of the general styles there were substyles that differed with respect to time spent.

In the re-test segment the 14-nominal analysis did not lead to the identification of information gathering styles. The three-interval analysis once again identified styles at two levels of generality. The general styles differed with respect to number of visits, and number of sources. One of the general styles contained three substyles that differed with respect to time.

8.2 PREPURCHASE INFORMATION GATHERING MODEL

The discussion of Areas of Study in Chapter 4 indicated three questions of concern with respect to modeling prepurchase information gathering. First, to what extent could major information gathering styles be identified? Second, what dimensions differentiated these styles? Third, if continuing work on information gathering styles and dimensions seemed useful, how should future research proceed? The next subsection discusses the styles and dimensions that were identified for the
furniture and appliance buyer samples. This is followed by a discussion of future research in this area. Finally, the marketing segmentation implications of information gathering styles are considered.

8.21 Information Gathering Taxonomy: Furniture Buyers

Analysis of data based on 287 furniture buyers led to the identification of seven information gathering styles. There were three dimensions differentiating the styles - number of store visits, number of sources, and time. There were three general styles differing in terms of number of store visits and number of sources. These styles were labeled thorough, moderately thorough, and non-thorough. There were also substyles differing in terms of time. They were labeled slow, moderate and fast. The styles formed the taxonomy indicated below.

```
FURNITURE BUYERS

THOROUGH

MODERATELY THOROUGH

SLOW (2) MODERATE (3) FAST (4)

NON-THOROUGH

SLOW (5) MODERATE (6) FAST (7)
```

In this taxonomy 22 percent of the subjects were not included because they were transients and could
not be categorized "correctly" in any one style. Of the categorized subjects five percent were thorough, 44 percent were moderately thorough, and 51 percent were non-thorough. At the thorough end of the style spectrum a group of 11 buyers used slightly more than three sources, made 20 store visits, and spent "years" considering the purchase. At the other end of the spectrum a group of 46 buyers used less than one source, made slightly more than two store visits, and considered the purchase for "days". The three-dimensional profile for each of the seven styles is presented below.

<table>
<thead>
<tr>
<th>STYLE</th>
<th>NUMBER SOURCES</th>
<th>NUMBER VISITS</th>
<th>TIME</th>
<th>NUMBER IN CLUSTER</th>
<th>PERCENT OF INTRANSIENTS</th>
<th>PERCENT OF TOTAL SAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) THOROUGH</td>
<td>3+</td>
<td>20</td>
<td>Years-</td>
<td>11</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>(2) SLOW</td>
<td>3+</td>
<td>3+</td>
<td>Months+</td>
<td>26</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>(3) MODERATE</td>
<td>3</td>
<td>8+</td>
<td>Months</td>
<td>31</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>(4) FAST</td>
<td>3½</td>
<td>6+</td>
<td>Weeks</td>
<td>42</td>
<td>19</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>99</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td>34</td>
</tr>
<tr>
<td>NON-THOROUGH</td>
<td>1</td>
<td>2+</td>
<td>Months</td>
<td>43</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td>(5) SLOW</td>
<td>1</td>
<td>2+</td>
<td>Months</td>
<td>43</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td>(6) MODERATE</td>
<td>1−</td>
<td>2−</td>
<td>Weeks-</td>
<td>25</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>(7) FAST</td>
<td>½</td>
<td></td>
<td>Days+</td>
<td>46</td>
<td>21</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>114</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>224</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td>78</td>
</tr>
</tbody>
</table>
The three-dimensional styles identified in this research corroborated the Katona and Mueller "non-deliberate" finding. Katona and Mueller (1955, page 53) found that approximately 25 percent of buyers were non-deliberate. Their measure of deliberateness was an index formed by summing across many aspects of information gathering. In this research 40 percent of buyers were in the non-thorough category (part of these, 15 percent, had considered the purchase for "months", and thus, would not have been non-deliberate in Katona-Mueller's terms). The style categorization also provided a view that was not evident in the Katona-Mueller findings, namely there was considerable variation in time within both the moderately thorough and non-thorough categories.

8.22 Information Gathering Taxonomy: Appliance Buyers

Analysis of data based on 259 appliance buyers led to the identification of five information gathering styles. There were three dimensions differentiating the styles - number of store visits, number of sources, and time. There were three general styles differing in terms of number of store visits and number of sources. These styles were labeled thorough/balanced, thorough/store intense, and non-thorough. There were also substyles differing in terms of time. They were labeled slow, moderate and fast. The styles formed the taxonomy.
indicated below.

APPLIANCE BUYERS

THOROUGH/
BALANCED
(1)

THOROUGH/
STORE INTENSE
(2)

NON-THOROUGH

SLOW
(3)
MODERATE
(4)
FAST
(5)

In this taxonomy 19 percent of the subjects were not included because they could not be categorized "correctly" in any one cluster. Of the categorized subjects 28 percent were relatively thorough, while 72 percent were non-thorough. The three-dimensional profile for each of the five styles is presented below.

<table>
<thead>
<tr>
<th>STYLE</th>
<th>NUMBER SOURCES</th>
<th>NUMBER VISITS</th>
<th>TIME</th>
<th>NUMBER IN CLUSTER</th>
<th>PERCENT OF INTRANSIENTS</th>
<th>PERCENT OF TOTAL SAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>THOROUGH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) BALANCED</td>
<td>4-</td>
<td>4½ Months-</td>
<td>41</td>
<td>19</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>(2) STORE INTENSE</td>
<td>2+</td>
<td>8 Months</td>
<td>18</td>
<td></td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>59</td>
<td>28</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>NON-THOROUGH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) SLOW</td>
<td>1½</td>
<td>1+ Months+</td>
<td>54</td>
<td>26</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>(4) MODERATE</td>
<td>1-</td>
<td>1+ Weeks</td>
<td>41</td>
<td>19</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>(5) FAST</td>
<td>1-</td>
<td>1+ Days</td>
<td>56</td>
<td>27</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>151</td>
<td>72</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>210</td>
<td>100</td>
<td>81</td>
<td></td>
</tr>
</tbody>
</table>
As with furniture purchasing a large proportion of the sample was very non-deliberate - 37 percent of the appliance sample used slightly less than one source, made slightly more than one store visit and considered the purchase for days or weeks. Once again the strength of the taxonomic approach was that it focused attention on differences within general groups. Within the relatively thorough buyer group some buyers primarily used stores as an information source, while others used a variety of sources. Within the non-thorough group there were considerable differences in time.

8.23 Comparison of Information Gathering by Furniture and Appliance Buyers

In both furniture and appliance taxonomies the three major dimensions differentiating the styles were number of sources, number of store visits, and time. In both taxonomies there were three styles differing primarily in terms of number of sources and number of store visits. The taxonomies were also similar in that each contained substyles differing with respect to time. Differences between the taxonomies were related primarily to the lack of thoroughness of appliance buyers. In the furniture taxonomy the non-thorough styles included 51 percent of the clustered subjects; they included 72 percent in the appliance taxonomy.
The non-thorough nature of appliance buyers was checked by comparing the two product samples on each dimension. Comparison of the response distributions for number of sources indicated that appliance buyers used fewer sources\(^2\). Comparison based on number of store visits and time indicated that appliance buyers made fewer visits and took less time\(^3\).

Explanation of why appliance buyers were less thorough was considered to be a step for future research. However, as an example of the potential causes, several variables external to the style analysis were also considered. The Fry-Portis survey gathered data on several variables useful in this regard. Measurements of the perceived product complexity indicated that product range for furniture products was greater than for appliance products. For example, 65 percent of the furniture buyers felt there were "big" differences in prices available, and 65 percent felt there were "big" differences in appearance. The corresponding percentages for appliance buyers were 35 and 28. The second measurement suggesting why appliance buyers were less thorough was the condition of the old product. Eight percent of furniture buyers reported the old one broken - 21 percent of appliance buyers. Finally, "shopping enjoyment" indicated 53

\(^2\) \(\chi^2\) was significant at \(\alpha\) of 0.10.
\(^3\) \(\chi^2\)'s were significant at \(\alpha\) of 0.01.
percent of furniture buyers enjoyed the shopping process "very much" - 35 percent of appliance buyers.

8.24 Significance of Information Gathering Taxonomies

Although statistical significance tests had not been developed, it was useful to estimate the significance of the information gathering taxonomies. Two questions were important. What could be concluded with respect to the styles and dimensions found? What could be concluded with respect to the areas where styles and dimensions were not found? Two major conclusions were possible with each question. The styles and dimensions that were found: (1) described important aspects of prepurchase information gathering, or (2) occurred by chance. The areas where styles and dimensions were not found: (1) were not important aspects of prepurchase information gathering, or (2) were poorly measured. These conclusions are described in the following diagram.

<table>
<thead>
<tr>
<th>FINDINGS</th>
<th>CONCLUSION</th>
</tr>
</thead>
<tbody>
<tr>
<td>POSITIVE</td>
<td>IMPORTANT DIMENSION</td>
</tr>
<tr>
<td></td>
<td>CHANCE</td>
</tr>
<tr>
<td>NEGATIVE</td>
<td>NOT IMPORTANT DIMENSION</td>
</tr>
<tr>
<td></td>
<td>POOR MEASURE</td>
</tr>
</tbody>
</table>

Thus, the significance was first assessed by considering the probability of the taxonomies occurring by chance. This probability was judged to be low for two
reasons. First, the styles identified had replicated across both halves of split-sample analyses. This suggested that the taxonomies were independent of the sample members. Second, the furniture and appliance taxonomies were similar in terms of the major styles and dimensions. It seemed unlikely that these similarities would occur unless the taxonomies were tapping important differences in behavior.

The second consideration was the probability that new measures of information handling would lead to the identification of further styles and dimensions. Since weaknesses in measurement were identified this probability was thought to be relatively high. In section 8.3 the measures used in this research are discussed in terms of their relative strengths for purposes of group identification. Changes in measurement are suggested. These changes are expected to lead to identification of further styles and dimensions.

In summary, the identified styles and dimensions appeared to describe important aspects of information gathering. New approaches to measurement, however, were expected to lead to additional styles and dimensions.

8.25 Modeling Prepurchase Information Gathering: Future Research

As discussed in subsection 8.23 comparison of
furniture and appliance buyers in terms of perceived product complexity, condition of old one, and shopping enjoyment suggested why furniture buyers were more thorough. In addition to explaining differences between product groups these factors would be expected to help explain differences in search styles. For example, "thorough" subjects would be expected to perceive the product as more complex than "non-thorough" subjects. Further, a broken old product would likely be associated with a "fast" search style.

Future research will relate formally external variables and the styles identified in this research. This is expected to provide an indication of the validity of the two taxonomies, and to increase understanding of prepurchase information gathering. Although data for assessing concurrent validity was available, this step was felt to be beyond the scope of this dissertation. One approach to this validation is multiple discriminant analysis. This approach could be used to relate prepurchase style to demographics, perceived product complexity, product experience, shopping enjoyment, condition of old product, etc.

The nature of the information gathering taxonomies led to two additional considerations for future research. First, the findings indicated the importance of considering the expected number of dimensions in terms of the
research objectives. Second, the areas of information gathering not included in the present taxonomies were considered important areas for future research. "Number of dimensions expected" is discussed next. "Areas not included in present taxonomies" is discussed in the section 8.3.

One objective of this research was to identify major\textsuperscript{4} information gathering styles. The two taxonomies that identified major styles were based on three relatively independent dimensions. A factor not obvious at the outset of the research was that major styles based on a large number of independent dimensions were unlikely. A simple example helps to illustrate this point. Suppose comparison of subjects on variables $X$ and $Y$ identifies three groups. Then if variable $Z$ is added, for the same

\begin{center}
\begin{tikzpicture}
    \draw[->] (0,0) -- (4,0) node[right] {$X$};
    \draw[->] (0,0) -- (0,4) node[above] {$Y$};
    \node (Gr1) at (1,3) [circle, draw] {Gr 1};
    \node (Gr2) at (1,1) [circle, draw] {Gr 2};
    \node (Gr3) at (3,1) [circle, draw] {Gr 3};
\end{tikzpicture}
\end{center}

groups to exist there must be a discriminant relationship between $Z$ and the 2-variable groups. An example of this

\textsuperscript{4} As defined in subsection 4.23 the term major was intended to mean a reasonably large proportion of the sample.
type of relationship is presented below. If a discriminant relationship does not exist, the third dimension will result in subdivision of groups. There is a finite probability of a discriminant relationship between each additional dimension and existing groups. Thus, as the number of independent dimensions increases, the probability of identifying major groups decreases.

Two implications for future research are indicated by this example. First, it is important to recognize the low probability of identifying major groups when intersubject comparisons are based on many independent dimensions. Thus, in future research it would be useful to limit the number of dimensions considered.

The second conclusion follows directly from the above. One approach to limiting the number of dimensions is to consider the areas of information gathering in a sequential manner. For example, one taxonomy could be based on major source dimensions and another on major type dimensions. Finally, the two taxonomies could be compared for similarities.
8.26 Information Gathering Styles and Market Segmentation

The potential of information gathering styles as an approach to market segmentation has implications for marketing managers. However, before discussing implications it is important to recognize that the description of styles provided by this research is not complete. Future research will attempt to identify factors underlying these styles. Since the styles identified indicate considerable differences in behavior, differences in underlying factors are also expected. As future research adds to the description of information gathering behavior, the implications for marketing managers are expected to become evident. However, the following example is presented to indicate the type of segmentation that might be expected.

Assume that an appliance manufacturer was considering the "thorough/store intense" style, and that he felt it represented a sufficient proportion of buyers to be of interest. Further research might suggest that buyers using this style had high perceived product complexity, low product knowledge and no immediate need. The conclusion might be that buyers were taking a long time in order to accumulate product information, and that the major source of information was retail outlets.

There are several types of marketing strategy
implications for a manufacturer. First, since the buyer is gathering information on a number of brands, product features are probably important to distinguish one brand from others. Second, since the buyer's major source of information is retail outlets, point of purchase material is particularly important. Third, the high perceived product complexity suggests that a major characteristic of the message, both point of purchase and salesforce, should be simplicity. Finally, it is not necessary to distribute through every outlet since the buyer visits many. However, the outlets used must provide a supportive, informed salesforce.

There are also several implications of the thorough/store intense style for the retailer. First, the product line carried should include a broad range of product features. Thus, after the buyer decides what he wants, he will return to the retail outlet with the broad line because the desired features will be available. Second, the retail salesforce should be supportive and provide information. Since the buyer is making many visits over an extended period, it is less important to make an immediate sale than to insure the buyer returns when he has made his decision. Third, the appropriate store image is one of friendly salespeople and a broad range of good products. This would influence
the nature of any newspaper advertising.

This example suggests the potential value of segmentation based on prepurchase information styles. Because of the speculative nature of this example, the assessment of information gathering styles remains incomplete. Confidence in the value of styles is gained by noting marketing management interest in this approach to segmentation. The manager of business research and forecasting for the Appliance and Television Business Group of General Electric was quoted in Business Week:

"Some people buy a refrigerator because they are moving into their first home. This buyer can take time to shop and compare, so feature differences are important. This is taken into account in formulating our ad program.

"Then there is the consumer who buys a refrigerator because the old one broke down. Because her need is more immediate she will be influenced most by point-of-sale advertising. And because her old refrigerator broke down, she will be especially concerned about quality, performance, and after-sales service. So to appeal to that customer, we give our advertising a different spin."5

This quote suggests that some marketing managers tailor marketing strategies to fit particular prepurchase styles. The information gathering taxonomies identified in this dissertation are seen as progress toward a formal model of information gathering styles.

5 Business Week, April 24, 1971, page 56
8.3 MEASUREMENT OF THE INFORMATION GATHERING PROCESS

The first aspect of information gathering measurement discussed is nature of the variables. Second, consideration is given to the relative strengths of variables used in this research. Finally, implications are suggested for the measurement of information gathering in future research. It is important to note that this discussion addresses measurement for research considering group-level models of information gathering. For example, other measures might be appropriate when modeling centers on individual behaviour.

8.3.1 Nature of Variables

Comparison of the findings based on the nominal and interval variables indicated the interval analyses to be superior in the identification of major styles. With nominal variables the only approach that led to style identification was the 14-variable exploratory analysis, and in this case the styles were of a general nature - deliberate and non-deliberate. With interval variables two-level taxonomies were identified in both the exploratory and re-test three-variable analyses.

The superiority of the interval approach was judged to be the result of two factors. First, the interval variables were more general than the nominal, and thus were a better match with the objective of identifying
major styles. The nominal variables measured many specific aspects of information gathering; the interval variables were more general in that they aggregated across several specific factors. Second, the interval format was stronger than the nominal. Nominal variables indicated whether or not a particular factor was part of the information gathering process. Interval variables indicated also the importance of the factor in the process.

As indicated in subsection 8.25, with an increasing number of independent dimensions the probability of discovering major groups decreases. Thus, for future research interested in major groups the options are to measure a limited number of general dimensions, or to measure many specific factors and aggregate these to form general dimensions. Aggregation could be based on judged similarity of specific factors or factor analytical techniques.

8.32 Strengths of Measures Used

Since the nominal approach to style identification did not prove fruitful, discussion of measurement concentrates on the interval variables tested. The conceptual model discussed at the outset of the dissertation is presented below. Based on this model, "number of features" and "number of alternatives" measured type of information. "Number of sources" and "number of store
visits" measured source. Time was measured directly. The

Conceptual Model of
Prepurchase Information Gathering

Primary Areas
Secondary Areas

TYPE

TOTAL
AMOUNT

SOURCE

INTENSITY

TIME

three areas of measurement (type, source, and time) are discussed next. The source variables are discussed first since "number of store visits" resulted in the clearest style differentiation. Consideration of store visits led to an understanding of the relative strength of this variable, and provided a useful guide when considering the other variables.

In the three-variable analyses subjects were differentiated by total store visits into three styles - many, moderate and few visits. These three general styles were found to be relatively distinctive. This finding indicated total store visits to be a strong variable. The relative strength of this variable was accounted for in two ways - total visits was an important dimension of behavior, and the approach was relatively strong. This discussion concentrates on the second aspect, strength of measurement.
The measurement strength of total store visits was attributed to two factors. The first factor was the nature of the phenomenon being measured. Since a store visit was a conscious, energy requiring activity, it was probably easily remembered as a specific aspect of information gathering. Thus, store visits was relatively easy to measure. The second characteristic which helped to account for its strength, was the wording of the questions. The subjects were asked which places of purchase they visited, and how many visits they made to each. The detail required by these questions was judged to have increased the accuracy of response by aiding recall.

The second source variable was number of sources. This variable was included in the three-variable analyses; however, the differences across search styles on this variable were not as great as they were for number of store visits, or time. It was useful to compare number of store visits and number of sources. Number of store visits measured the relative importance to buyers of a particular source. On the other hand, number of sources was a count of sources considered. It was concluded that the latter was a weaker measure of behavior since

---

6 Question wording appears in Table 5.5
it aggregated across a variety of sources, and since the relative use of each source was not measured. That is, the generality of number of sources did not provide meaningful discrimination across subjects. This suggested an important balance in future attempts to identify major groups. The number of independent dimensions considered must be limited, and yet dimensions too general to provide meaningful differences must be avoided. Recommendations for the measurement of source dimensions are presented later in this chapter.

The generality of the time measure was consistent with the identification of general information gathering styles. However, the appliance data caused concern since recent purchasers reported longer time spent on decision making than did non-recent purchasers\textsuperscript{7}. Under-reporting of non-recent purchasers, due to forgetting, was judged to be the probable cause of the difference. Thus, it was important to consider steps to increase the accuracy of the time measurement. Experience with number of store visits suggested that increasing the detail required by the question could prove useful. For example, questioning could begin in a relatively open-

\textsuperscript{7} See Table 7.6
ended manner. This could then be followed by probes attempting to help the respondent recall when she first started thinking or talking about the purchase. Major events, such as "before Christmas" and "after summer holidays", could be used to help her recall. Another approach to probing recall would be to see if the respondent could remember when she started various types of information gathering, such as first looked at the product in stores, or first talked to husband about the product.

The final area measured was type of information gathered. Neither number of features nor number of alternatives was included in the 3-variable analyses. One weakness in the features measurement was suggested - the wording of the question was suspected of causing response bias. As a result of style identification analysis, conclusions with respect to other weaknesses were also apparent. First, it seemed probable that number of features was too general a measure. That is, aggregation across specific features reduced meaningful discrimination across subjects. Second, the four nominal features variables (price, quality, style and size) were judged to be an appropriate level of generality. However, these measures could be strengthened by determining the extent to which feature was considered.

Alternatives as measured in the Fry-Portis survey
did not prove useful in identifying groups via taxonomic procedures. Future attempts to measure type of information considered would be concentrated on measures similar to the features variables. Future measures of type are discussed in the next subsection.

8.33 Implications for Future Measures of Information Gathering

Each area of information gathering (types, sources, and time) merits special attention in future research. Recommendations are made for measuring each area.

Intuitively there would seem to be a number of basic type-dimensions underlying the array of product characteristics. For example, a desire for long trouble-free operation may underlie responses such as durability, quality, confidence, convenience, brand, and service. An interest in succinct description of information types leads to future research in this area. The approach recommended for tapping basic type-dimensions is to provide the respondent with a list of product characteristics, and to have her rate the importance of each on a multi-point scale. (The responses to the Fry-Portis open-ended features questions would provide a useful characteristics list.) Factor analysis could then be used to attempt the identification of major type-dimensions in the features data. If major dimensions are identified, the relative importance of each dimension
to each subject could be determined by calculating factor scores. Finally, group identification could be based on the relative location of subjects in the reduced-space.

A similar approach is recommended for measuring source usage. That is, develop an extensive list of sources; have each respondent rate the importance of each source; factor analyse the source ratings to identify major source dimensions; compare subjects across these dimensions. An alternative approach is to take direct similarities measures for a list of sources. These similarity measures could be analysed using nonmetric multidimensional scaling. The source configuration would provide an indication of the major source dimensions.

The recommended approach to measuring time is to attempt aiding recall by means of the open-ended probing suggested in the subsection 8.32. This approach is expected to provide a measure of time that would be useful for group identification. Further conceptual and empirical work that attempts to describe more fully the time aspects of information gathering would also be useful. One possibility is to consider time-source stages. For example, Stage I: early in the process she discusses product with friends. Stage II: in the middle of the process she notices store displays during the course of other shopping. Stage III: late in the
process she makes specific trips to see retail displays.

8.4 NUMERICAL TAXONOMIC PROCEDURES

The research effort that centered on numerical taxonomy had as an objective the selection of set of procedures to be used for identifying buyer search styles. This included consideration of available procedures, and the development of new ones. Three factors guided the selection. First, the procedures had to be suited for use with the Fry-Portis measures of prepurchase information gathering. Second, it was desirable for the procedures to focus on major groups; categorization of every individual subject was less important. Third, to increase confidence in the clustering results it was desirable for the procedures to indicate the extent to which groups differed, and the extent to which groups replicated across samples,

Numerical taxonomic procedures are discussed in three parts. First, general conclusions regarding the strengths and weaknesses in this application are discussed. Second, consideration is given to specific aspects of this application. Finally, future applications of numerical taxonomic procedures are suggested.

8.41 General Merits in this Application

Two major strengths of numerical taxonomic procedures were evident in this research. First, these procedures
facilitated the exploration of a large set of data with a minimum of prestructuring. This exploration led to the identification of major behavioral styles and dimensions. The exploration also provided an approach to assessing information gathering measurements. Second, the use of taxonomic procedures resulted in the comparison of groups based on a profile of variables. Because of this profile approach, it was possible to identify subgroups within general groups. In particular there were subgroups that were similar in terms of store visits and number of sources but different in terms of time. This type of interaction between descriptor variables would not have been evident using a single variable or summed index model.

The primary drawback of these procedures was the degree of commitment and effort required to use them fruitfully. Although the procedures were conceptually very simple, the weakness with respect to descriptive and inferential statistics necessitated a high degree of judgment be applied to the meaning of the results. To make the required judgments the researcher had to appreciate the relative merits of a variety of procedures. First, there were a number of possible measures of proximity. Joyce and Charron (1966) discussed nine similarity measures; Green and Rao (1969) considered ten proximity measures. Second, there were many possible
clustering algorithms - Carlson (1970) reviewed 26. Third, considerable effort was required to evaluate clustering results. Appendix B indicates the sequence that was followed in an attempt to insure meaningful clustering results. The approximate time required for one sequence was 12 to 14 hours. This was after the sequence was developed with computer programs written and did not include turn around time for batch processing.

In summary, it was difficult to evaluate the cost and value of numerical taxonomic procedures. Although the costs in terms of research effort were relatively well defined, the value of information gathering taxonomies remained to be determined by future research. However, because of the potential for unstructured identification of natural groupings and the advantage of profile information, the value was judged to outweigh the costs.

8.42 Merits of Specific Procedures

This subsection deals with the merits of the cluster identification procedures developed as part of this research - reverse matrix clustering and overlap assessment. Reverse matrix clustering was developed to identify the influence of ties in the proximity matrix. The procedure involved clustering each proximity matrix twice. After the first clustering the subject-order in
the matrix was reversed, and the matrix was reclustered. Subjects not remaining in the same cluster in both clusterings, were the subjects influenced by ties. Use of this procedure led to the recognition that identifying subjects influenced by ties was in effect identifying subjects that could not be linked correctly to any one cluster. These subjects were termed transients. The removal of transient subjects from the cluster set focussed attention on concentrations in the subject-scatter.

A major question to be resolved by further research is to determine the benefit of clustering a proximity matrix more than twice. To review, given the presence of ties in a proximity matrix, cluster formation was to some degree a function of the matrix subject-order. Using the diameter method, the "next" cluster in the hierarchical scheme was found by a systematic scan of the proximity matrix to find the "next most compact" cluster. During the scan if two clusters of the same compactness were found (a tie), the first cluster found formed the "next" cluster. Because the scanning procedure was fixed, the matrix subject-order influenced which cluster was formed next. Since multi-way ties were possible, and since there were n-factorial possible subject-orders, clustering the matrix three or more times was considered.
The objective of multi-order matrix clustering would be the identification of completely distinctive clusters by eliminating all possible transients. Since overlap assessment was used to indicate cluster distinctiveness, it was judged that the identification of all transients was not critical. That is, if reverse matrix clustering failed to identify a group of transient subjects between a pair of clusters, the result would be greater overlap of these two clusters. Thus, overlap assessment would provide an indication of unidentified transients. Since reverse matrix clustering provided complete change in subject-order, and preliminary tests on synthetic data had indicated satisfactory results, multi-order matrix clustering was not pursued. Further research is required to determine the merits of multi-order clustering.

The second procedure developed for evaluating clustering results was overlap assessment. With a large number of subjects, or more than three variables, results could not be plotted, and thus understanding the distinctiveness of the clusters was difficult. Without a plot, replication comparison for clusters from split-sample analyses also was difficult. To provide a guide to the distinctiveness and replication, overlap assessment was developed. This procedure involved three steps - first, scaling cluster centroids using nonmetric multidimensional scaling - second, plotting the scaled centroid configura-
tion - and third, plotting at each centroid the appropriate cluster diameter. This proved to be a useful guide to relative cluster location, and relative cluster size.

During the development of overlap assessment, several sets of "known" data were analysed and the procedure performed well. However, further research on a greater range of test data would be useful. For example, data combinations should be tested that differed in terms of number of variables, degree of intercorrelation among variables, format of the variables, and presence of noise in the data.

8.43 Future Applications in Marketing

The final consideration with respect to numerical taxonomy is to discuss future applications in marketing. Two factors are central to the use of these procedures - the purpose of the research, and the research effort available. The major purpose of these procedures is the exploratory classification. In applications similar to this research the potential for identifying underlying types and dimensions is particularly useful.

The predominant drawback of numerical taxonomic procedures is the considerable research effort required to evaluate the clustering results. In this dissertation three steps were taken to attempt to insure meaningful results. First, two new procedures were developed,
reverse matrix clustering and overlap assessment; second, split-sample comparisons were included; third, the analyses was done in two segments - exploration and re-test. For the most part similar precautions will be required in future marketing research applications of numerical taxonomy.

There is, however, one type of application that simplifies the use of taxonomic procedures - as complimentary analysis to multidimensional scaling (Green and Carmone, 1967). Since in this application similarity is measured directly and the stimuli configuration is plotted, two major problems associated with the use of numerical taxonomy are minimized. That is, the selection of appropriate proximity matrix is eliminated, and interpretation of clusters is simplified by virtue of the brand configuration plot.
REFERENCES


Bruner, J.S., J.J. Goodnow, G.A. Austin (1959), A Study of Thinking, John Wiley and Sons.


Cox, Donald F., ed. (1967), Risk Taking and Information Handling in Consumer Behavior, Harvard University.


March, James G., and Herbert A. Simon (1964), Organizations, John Wiley and Sons.


McNemar, Quinn (1962), Psychological Statistics, John Wiley and Sons.


APPENDIX A

REVERSE MATRIX CLUSTERING:
IDENTIFICATION OF TRANSIENT CLUSTER MEMBERS

The system developed to identify transient cluster members is shown in Table A1. Three steps are involved. First, the major clusters in the forward clustering are identified and numbered. This is shown as "Step One". The first split in the hierarchy divides the subjects into two clusters - these are numbered "1" and "2" at Level A. Each of the first two clusters are subdivided at Level B. There is one further subdivision at Level C. There are five clusters in the hierarchy at Level C. The first cluster on the left is given the identifying number "11". This indicates that this group of subjects is in cluster 1 at Level A and cluster 1 at Level B. The other clusters from left to right are identified 12, 21, 221, and 222.

The second step is to write the forward clustering identifiers above each subject in the reverse clustering output. This is shown as "Step Two". Thus subject 16 (on the left side of the reverse clustering) has "222" above it, indicating its cluster membership at various levels in the forward clustering.
The third step is to identify major clusters in the reverse clustering, and then compare forward and reverse cluster membership. In reverse clustering the first split is between subjects 12 and 11. By scanning the cluster identifiers it can be seen that the subjects in the left cluster (16, 19, 15, 14, 20, 18, 17, 13, 12) were in cluster 2 at Level A. For the most part subjects in the right hand cluster were in cluster 1 at Level A - two subjects (11 and 10) were in cluster 2. Because these two subjects "changed clusters" (transients) at the 2-cluster level an asterisk (*) is placed above them. This is shown as "Step Three". The transients for Levels B and C are identified in a similar manner.
### TABLE A1

### FORWARD CLUSTERING

<table>
<thead>
<tr>
<th>LEVEL</th>
<th>STEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>ONE</td>
</tr>
<tr>
<td>B</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 2 6 7 1 3 4 8 9</td>
<td>0 1 2 4 3 7 6 5 9 8 0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>.. xxx . xxx xxx xxx . xxx . xxx xxx</td>
</tr>
<tr>
<td>0.14</td>
<td>. xxxxxx xxxxxx xxx xxx xxxxxx xxx</td>
</tr>
<tr>
<td>0.20</td>
<td>xxxxxxx xxxxxxx xxx xxx xxxxxx xxx</td>
</tr>
<tr>
<td>0.28</td>
<td>xxxxxxxxx xxxxxxxxx xxx xxxxxxxxxx</td>
</tr>
<tr>
<td>0.40</td>
<td>xxxxxxxxxxx xxxxxxxxxxx xxxxxxxxxxxxx</td>
</tr>
<tr>
<td>0.50</td>
<td>xxxxxxxxxxxxxxxxxxx</td>
</tr>
<tr>
<td>0.90</td>
<td>xxxxxxxxxxxxxxxxxxxxxxxxxx</td>
</tr>
</tbody>
</table>

### REVERSE CLUSTERING

<table>
<thead>
<tr>
<th>LEVEL</th>
<th>STEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>TWO</td>
</tr>
<tr>
<td>B</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>*</th>
<th>*</th>
<th>*</th>
<th>*</th>
<th>*</th>
</tr>
</thead>
</table>

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2 2 2 2 1 2</td>
<td>2 2 1 1 1 1</td>
<td>1 1 1 1 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 2 2 2 2 2</td>
<td>2 2 1 2 2 2</td>
<td>1 1 1 2 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 2 2 1 2 2</td>
<td>2 2 1 1 1 1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Level</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>.. xxx . xxx xxx xxx . xxx . xxx xxx</td>
</tr>
<tr>
<td>0.14</td>
<td>. xxxxxx xxxxxx xxx xxx xxxxxx xxx</td>
</tr>
<tr>
<td>0.20</td>
<td>xxxxxxx xxxxxxx xxx xxx xxxxxx xxx</td>
</tr>
<tr>
<td>0.28</td>
<td>xxxxxxxxx xxxxxxxxx xxx xxxxxxxxxx</td>
</tr>
<tr>
<td>0.40</td>
<td>xxxxxxxxxxx xxxxxxxxxxx xxxxxxxxxxxxx</td>
</tr>
<tr>
<td>0.50</td>
<td>xxxxxxxxxxxxxxxxxxx xxxxxxxxxxxxx</td>
</tr>
<tr>
<td>0.90</td>
<td>xxxxxxxxxxxxxxxxxxxxxxxxxx</td>
</tr>
</tbody>
</table>

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>

| 1 1 1 1 2 1 1 1 1 1 1 | 1 1 1 1 1 1 1 |
| 6 9 5 4 0 8 7 3 2 1 0 | 9 7 8 4 5 6 2 3 1 |
APPENDIX B

STANDARD ANALYSIS PROCESS

For purposes of describing the analysis process assume that subjects were being compared on three interval descriptor variables (X, Y and Z). The following steps would be taken.

1. For Group A of split-sample standardize the variables to zero mean and unit variance

   \[ x_i = (X_i - \bar{X}) / S_x \]

   where \( x_i \) is the standardized score on variable X for subject i; \( X_i \) is the raw score on variable X for subject i; \( \bar{X} \) and \( S_x \) are the mean and standard deviation of variable X across subjects.

2. Calculate euclidean distance matrix

   \[ D_{ij} = \left[ (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 \right]^{1/2} \]

   where \( D_{ij} \) is the euclidean distance between subjects i and j; \( x_i, y_i, \) and \( z_i \) are the standard scores for subject i on the three variables.

   The distance matrices were calculated using a Fortran IV program - PROX. This program was written for this research and was run on an IBM 1130 computer.

3. Analyze distance matrix using modified "diameter" algorithm. Two features were added to Johnson's clustering algorithm. First, average distance between and within clusters was calculated for each link in the hierarchical scheme. Second,
after clustering the input distance matrix, the subject-order of the matrix was reversed and the matrix reclustered. Both forward and reverse hierarchical clusterings were provided as output. The modified "diameter" algorithm was run on an IBM 7040, and could cluster up to 150 subjects.

4. Identify intransient clusters by comparing forward and reverse hierarchical schemes. This comparison is discussed in Appendix A.

5. Identify "assessment" clusters. Clusters of interest were selected based on intransiency, between/within statistics and cluster diameters.

6. Calculate cluster centroid for each assessment cluster. Cluster centroids were calculated using PROFIL, a Fortran IV program written for this research and run on an IBM 1130.

7. Repeat steps 1 to 6 on Group B of split-sample.

8. Calculate inter-centroid distance matrix for centroids from Group A and Group B assessment clusters. PROX was used to calculate this distance matrix.

9. Analyse centroid distance matrix using nonmetric multidimensional scaling. The TORSCA program (Young, 1968) was used for this step.

10. Plot centroid configuration in appropriate dimensionality.
11. Scale and plot cluster diameters.