

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

June 1973

PROFIT ORIENTED DATA ANALYSIS FOR MARKET
SEGMENTATION: AN ALTERNATIVE TO AID

Working Paper No. 77

by

Claude R. Martin, Jr.

Associate Professor of Marketing

and

Roger L. Wright

Associate Professor of Statistics

FOR DISCUSSION PURPOSES ONLY

None of this material is to be
quoted or reproduced without the
express permission of the Division
of Research.

ABSTRACT

The authors report on a new AID-like algorithm that overcomes some of the conceptual problems involved in using AID for consumer segmentation. Using a simple cost-profit formulation they segment 356 women consumers on the basis of 26 dimensions that are consistent with the Howard-Sheth hypothetical construct.

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

INTRODUCTION

Marketing researchers are discovering that a relatively new analytical tool, Automatic Interaction Detector (AID),^{1/} can be a powerful tool for examining consumer behavior data because of its versatility in revealing complex interactive associations and the ease of interpreting its analysis as market segmentation.^{2/} This versatility becomes particularly clear in the analysis of some dichotomous dependent variables. There is some doubt, however, whether the division of respondents on the basis of AID's least-squares algorithm brings a segmentation that meets the profit objectives of most marketers. This paper presents a description of a new AID-like algorithm (Survey Implemented Market Segmentation--SIMS) that offers an alternative to AID and is derived from a simple cost-profit formulation.

SURVEY-DATA INPUT

Two inputs, a battery of in-depth interviews and the behavioral models developed by Howard and Sheth,^{3/} were used to formulate the ultimate research instrument, a mail questionnaire. The first input consisted of 75 interviews with women consumers in three southwestern Missouri stores. The women were approached as they completed an apparel purchase and were interviewed concerning that purchase decision. The results were then correlated with the Howard-Sheth modeling concepts to derive the mail questionnaire. The variables tested were categorized as follows: behavior, predispositions, information and product cues, demographics, and buyer goals. Each woman was asked to retrace her most recent purchase of an item of apparel for herself, and she was questioned about specific variables (Figure 1) as they related to that purchase.

Two retail-trade areas in Missouri--Joplin and Springfield--were chosen for this study. There were several reasons for selecting these areas: (1)

Demographics:

Marital status
Age
Employment status of respondent
Employment status of husband of
respondent, if married
Number and ages of children
City of residence

Predispositions:

Negative colors--garment colors re-
spondent would not buy
Negative fabric characteristics --
fabrics respondent would not buy
Garment care characteristics wanted
Wardrobe accessory matching
Upper and lower price limits to
purchase
Had charge account where shopping
and buying reported
Previously bought apparel in store
of purchase
Prepurchase planning:
General
Specific--positive color wanted
positive fabric wanted

Product and Information Cues:

Comparison shopping at alternate
stores
Utilization of price limitations
Method of payment
Sought out particular sales clerk
Use of "shopping pals"
Used sales clerk evaluations of style
and fit of garment
Evaluation of mass media helpfulness in
purchase decision

Buyer's Goals:

Self-evaluation of fashion awareness
Factors used in developing level of
fashion awareness
Shopping enjoyment in buying clothes
for self

Behavior:

Coordinating items purchased
Type of garment purchased
Number of stores shopped
Number of stores shopped on day of
purchase
Color of garment purchased
Fabric of garment purchased
Garment care requirement for item
purchased

The local merchants agreed to cooperate; (2) The two areas showed differences in socioeconomic status and growth; ^{4/} and (3) The two areas were geographically close, thereby controlling regional differences. The major factors differentiating the two markets were population growth, educational levels attained, median and mean income levels, and median value of housing (Table 1).

The mail survey obtained 356 usable responses. The distribution of the respondents was compared to the age, marital status, and employment distributions of the 1970 census to check for a nonrepresentative sample, and it was found that the distribution of the respondents was similar in configuration to that of the general population.

The original in-depth interviews led to the conclusion that there are differences in consumer behavior associated with different types of retail stores. The respondents were asked to identify the store in which their last purchase of an item of apparel had been made. With the assistance of a five-member retailer panel in both cities, the 96 stores selected were classified into general categories. Two of these categories were selected for further analysis by both the AID and SIMS algorithms: women who bought in independent department stores and women buying in women's specialty stores. The independent department store is defined as that store which offers a broad line of merchandise but is not controlled by any of the so-called "big three"--Sears, Wards, or Pennys. The more obvious classification of women's specialty stores consists of those merchants who primarily sell women's clothing.

AID ANALYSES

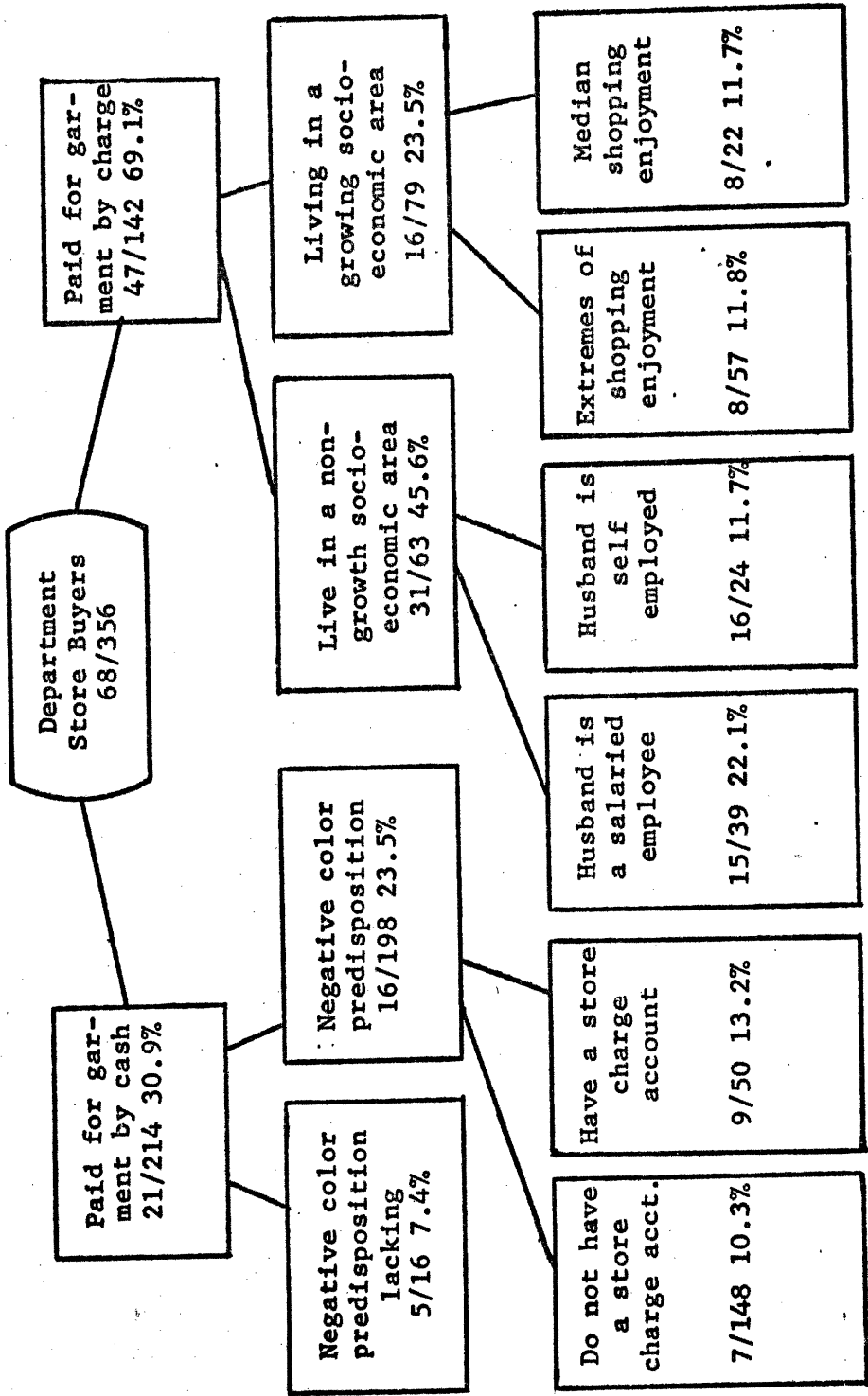
The results from the use of the AID algorithm are seen in the tree diagrams in Figures 2 and 3. The ratio on the left side of each box is the number of buyers for that type of store to the total number in the sample at that point on the tree. The percentage on the right is the share of all buyers (e.g., in Figure 2, the share of all department-store buyers) occupying

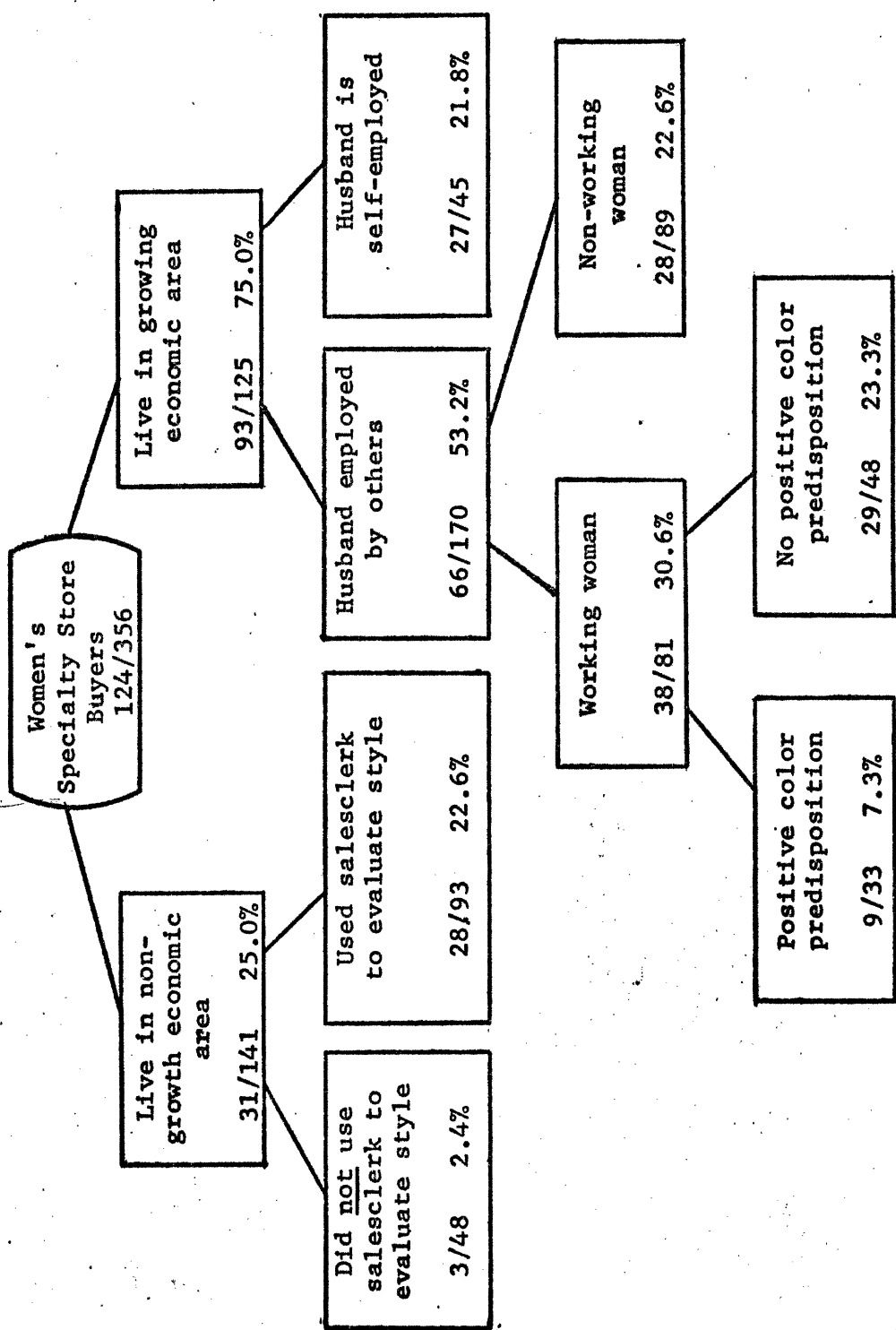
TABLE 1

1970 Census Characteristics of Joplin and Springfield, Missouri

Census Characteristic	Springfield Standard Metropolitan Statistical Area (SMSA)	Joplin Retail Trade Area	
		Jasper County	Newton County
Population growth (in percentage)	8.5	6.9	8.1
Median years of education for males	12.3 years	11.8 years	11.5 years
Percentage of population having completed high school	58.9	48.7	47.3
Median income	\$ 8,215	\$7,312	\$6,887
Mean income	\$ 9,310	\$8,410	\$7,785
Owner-occupied house- hold's median value	\$13,900	\$9,000	\$9,800
Renter-occupied house- hold's median rent	\$73	\$55	\$55

Source: Bureau of the Census, Census of the Population, 1970, PC(1), A27, B27, C27 (Washington, D.C.: Government Printing Office, 1971).





that particular branch on the tree. The tree diagrams show graphically the characteristics of buyers of women's ready-to-wear clothing in department stores (Figure 2) and in women's specialty stores (Figure 3). These diagrams should be useful to the decision maker in identifying major market segments.

MARKET SEGMENTATION BY AID: A CRITIQUE

A major point can be raised about this use of AID: Are the resulting groups consistent with the marketer's interests in segmentation? A look at AID's criteria offers some answers. Similar to stepwise linear regression programs, AID searches at each step for the single predictor that best explains the variance of the dependent variable. Like one-way analysis of variance, the criteria is the weighted sum-of-squares between the means of the two subgroups formed. This procedure is equivalent to minimizing the sum-of-squares within the resulting two subgroups. Taking advantage of the dichotomous nature of our dependent variable, the within sum-of-squares may be expressed simply as follows:

$$\sum_{i=1}^2 s_i (t_i - s_i) / t_i$$

where s_i equals the number of buyers in subgroup i , and t_i equals the total number of respondents in subgroup i . AID chooses its segments to minimize this equation.

Another important aspect of the AID algorithm is the criteria used to select the next group for further splitting. Continuing to apply regression concepts, AID chooses the group with the largest within sum-of-squares.

The proposition advanced here is that in certain circumstances financial considerations can provide an even more relevant basis for market segmentation analysis than can AID's basis of stepwise regression and analysis of variance concepts. A note of caution, however, is necessary: not all of the many factors ultimately used in formulating a marketing strategy are accounted

for in any algorithm, but there is an algorithm that is better for handling some of the relevant dimensions.

MARKET SEGMENTATION BY SIMS: A DIFFERENT RATIONALE

Other factors being constant, the potential of a particular group depends on the group size, t , and on the number of buyers in a group, s . As a first formulation, two parameters seem to be involved--the marginal contribution to overhead and profit of selling to a buyer (b) and the marginal cost of marketing to a nonbuyer (c). The net contribution to overhead and profit in marketing to a group of size t with s buyers is as follows:

$$sb - (t-s)c$$

Or this may be rewritten as:

$$t(b+c)(s/t-r)$$

where r denotes $c/(b+c)$. We call r the trade-off factor; it reflects the increase in the number of buyers necessary to offset a unit increase in the group.

We assume that the contribution to overhead and profit of marketing to any group does not depend on marketing to other groups and that the profit in not marketing to other groups is zero. Then, clearly, total profit is maximized by marketing to all groups for which s/t exceeds r . We recognize that this formulation represents an oversimplification of the actual marketing and economic considerations; for example, in attempting to simplify we have purposely not built in the complexity of fixed variable costs. We feel, however, that this formulation provides a natural basis for initial statistical analysis without imposing undue demands on the analyst.

Consider a hypothetical independent variable X taking several different values. Each of these values determines a corresponding subgroup of respondents, as in Table 2. The strategy is to market in those subgroups that have their proportion of buyers greater than r , that is with $s_i/t_i > r$. This strategy

TABLE 2

Illustrative SIMS Analysis

Values of Variable X i	Total Number of Cases t_i	Number of Buyers s_i	Proportion of Buyers s_i/t_i	Marketable?	Contribution to Profit (in dollars)
0	163	9	.055	No	-\$127.00
1	14	1	.071	No	- 10.00
2	45	10	.227	No	- 5.00
3	90	32	.356	Yes	+ 38.00
4	44	16	.363	Yes	+ 20.00
TOTAL	356	68		No	- 84.00

suggests that we split the original group into two segments--the first comprised of all marketable subgroups and the second of the rest of the groups. Assume that the cost (c) of marketing to a nonbuyer is assessed at \$1.00 and that the contribution to profit and overhead of marketing to a buyer (b) is set at \$3.00, so that r equals .25. In Table 2, the marketable segment formed by the variable X comprises 134 cases, of which there are 48 buyers. The remaining 222 cases, including 20 buyers, fall into the nonmarketable segment. Then the negative contribution (loss) from marketing to all five segments is \$84.00, compared to a profit maximization of \$58.00 for the marketable segment alone.

This strategy, of course, depends on the choice of the independent variable X. Another variable will define a different set of subgroups and result in a different segmentation. Clearly the best single independent variable is the one that maximizes the profit of marketing in its marketable segment. The SIMS program considers each available variable, forms the corresponding segmentation into the marketable and nonmarketable segments, and calculates the profit of the marketable segment. The original group is split into two segments using the variable that maximizes the profit of its marketable segment.

After the split of the original sample into two segments, of which one is marketable, one of these segments will be selected as a candidate for further segmentation. Although we cannot with certainty anticipate the contribution to profit and overhead of further segmentation, the contribution of a perfectly discriminating segmentation may be easily evaluated. Consider a group of size t with s buyers. If $s/t > r$, then this group is scheduled for marketing so that the contribution to overhead and profit of perfect further segmentation is the elimination of the cost $(t-s)c$ of marketing to the nonbuyers in the group. On the other hand, if $s/t < r$,

then this group as a whole is nonmarketable so that the contribution of perfect further segmentation is sb, that is, the contribution of marketing to only the buyers within the group. In this way we define the potential profit of further segmentation for each group. The group with the greatest potential is selected as a candidate for further segmentation, and its own segmentation is determined exactly as we have described above. By iteration, a sequence of segmentations analogous to AID is produced.

Suppose, for example, that variable X considered in Table 2 produces the best initial segmentation. Then either the marketable segment formed or the nonmarketable segment is selected for further segmentation. The potential of the marketable segment is \$86.00 and for the nonmarketable group only \$60.00, so that the marketable segment would be selected next. It may sometimes be desirable to require that a group's potential exceed some fixed level before being considered eligible for further segmentation. This requirement will eliminate from further consideration groups with very small numbers of buyers, nonbuyers, or both. The result will be the earlier termination of the recursive analysis.

OPERATING DATA INPUT

In the southwestern Missouri project the actual annual operating data from 21 cooperating stores was used. This data from the women's ready-to-wear departments in the two types of stores analyzed is shown in Table 3.

For the analysis of the market for independent department stores, the applicable figures are $c = \$.89$, and $b = \$10.20 - \$3.97 = \$6.23$; for women's specialty stores, $c = \$1.76$, and $b = \$11.21 - \$6.44 = \$4.77$.

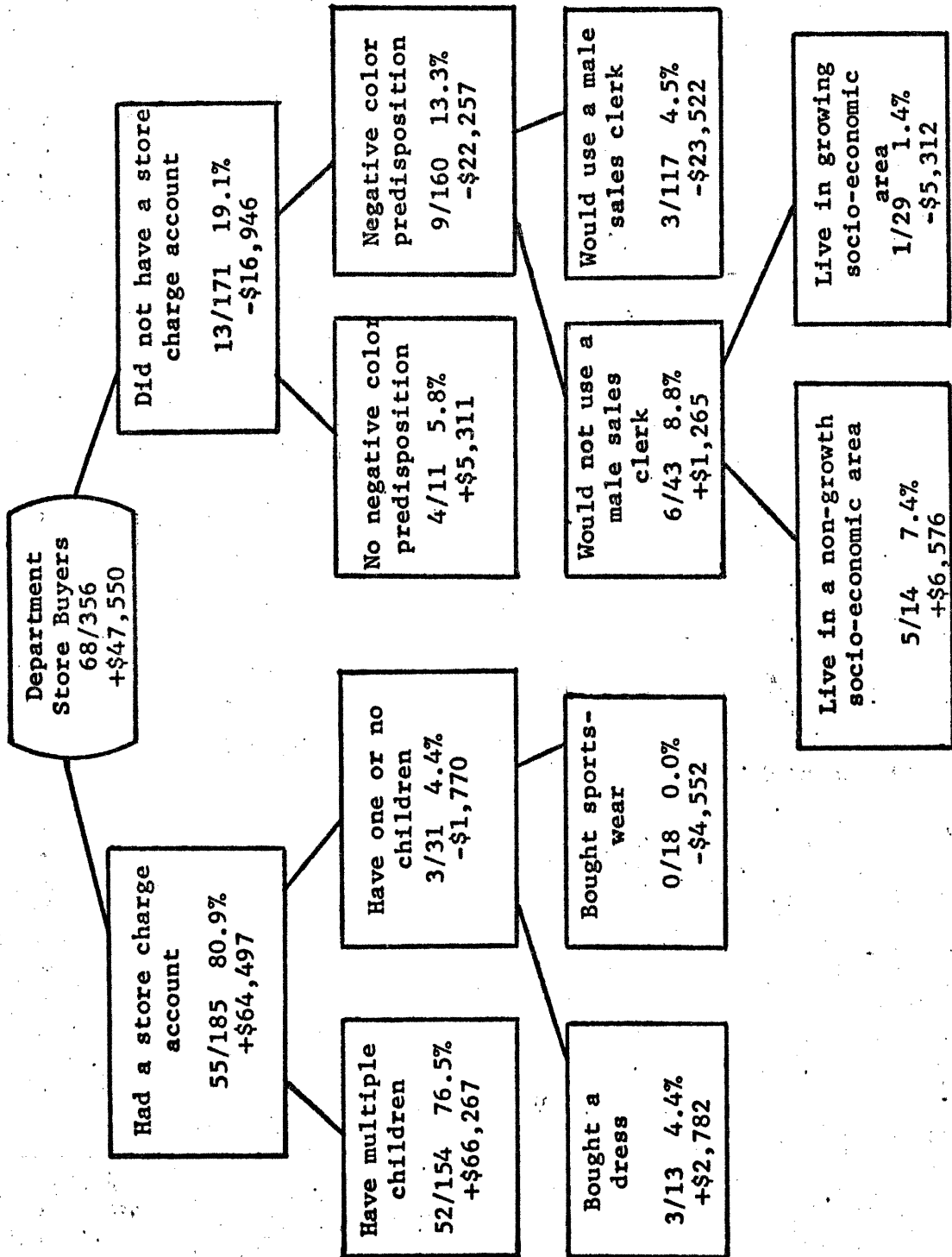
SIMS RESULTS

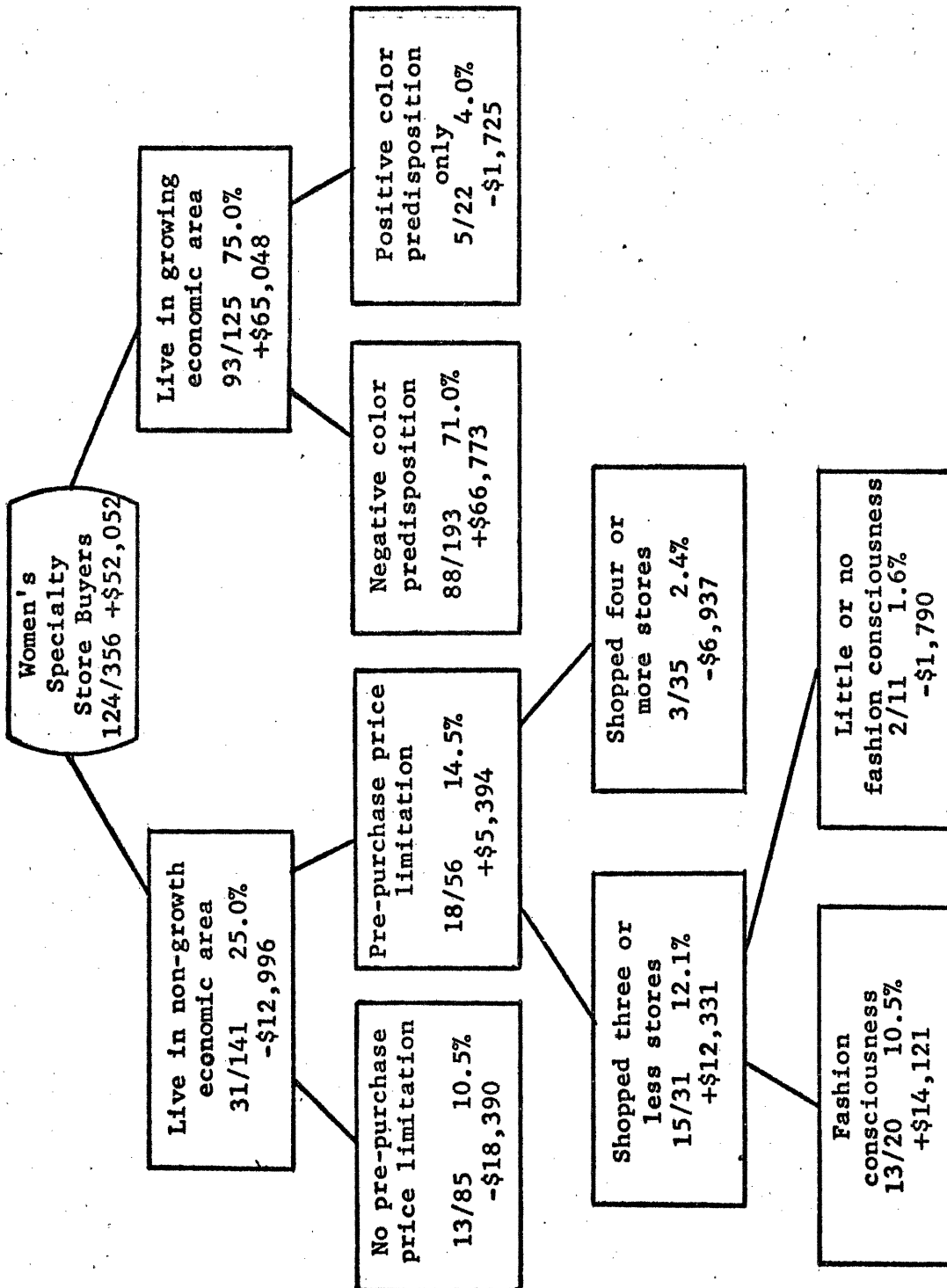
The tree diagrams in Figures 4 and 5 show the results after the data were subjected to the SIMS algorithm. The ratio on the left side of each

TABLE 3

Operating Data for Women's Ready-to-Wear Departments

Item	Independent Department Stores	Women's Specialty Stores
Average net sale	\$25.30	\$27.50
Cost of goods sold	\$15.48	\$16.83
Gross margin	\$ 9.82	\$10.67
Account receivable revenue	\$.38	\$.54
Adjusted gross margin	\$10.20	\$11.21
Variable costs assignable only to a buyer	\$ 3.08	\$ 4.68
Survey ratio of buyers to total sample	.191	.348
Joint variable costs assignable to both buyers and nonbuyers on basis of their ratio (c)	\$.89	\$ 1.76





box is that of buyers to the total in the sample at that same point on the analytical tree. The percentage on the right side is the share of all buyers occupying that particular segment of the analytical tree.

We have assumed in this study that the sampling of women fashion consumers in the two markets was representative. In those markets, the 1970 census data show a total of 101,171 women over 18 years of age--a ratio of 284.1882 women in the population to each woman in the sample. In the SIMS program we multiplied the contribution figure for each segment by this ratio in order to produce our estimate for the combined marketing areas. Therefore, the dollar figure found in each box is the estimated contribution to overhead and profit produced for all stores by that segment.

Based on our previously stated assumption of no interdependence between segments, the independent department stores' marketers could sell to only those on the left-hand side of the tree and change the marginal contribution from \$47,550 to \$64,497. Thus the stores would have to face up to dropping 19.1 percent of their market share but with the right-hand segment's corresponding net cost (negative contribution) of \$16,946. The same type of assessment can be made of the women's specialty store buyers: selling to only the right-hand branch of the tree would change the marginal contribution from \$52,052 to \$65,048, with a corresponding drop in market share of 25.0 percent.

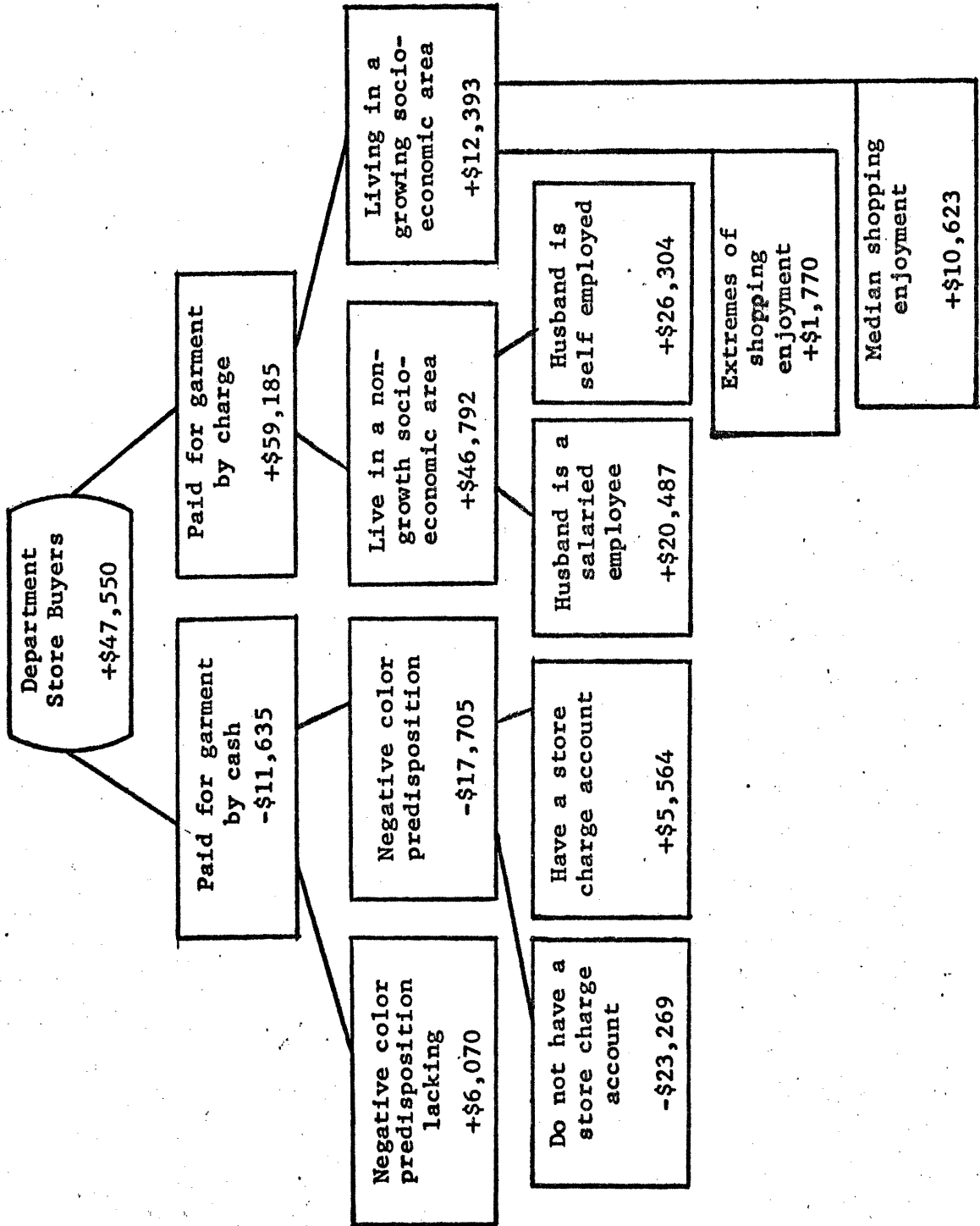
Each of the store types could maximize its contribution to profits and overhead by concentrating on only those positive branches of each SIMS tree. In the case of the independent department store buyers this contribution amounts to +\$80,936; for women's specialty store buyers it is +\$80,894. We recognize that there may be instances where differentiating between market segments is operationally impossible--e.g., could a women's specialty store sell only to those who have a negative color predisposition and not appeal to those with

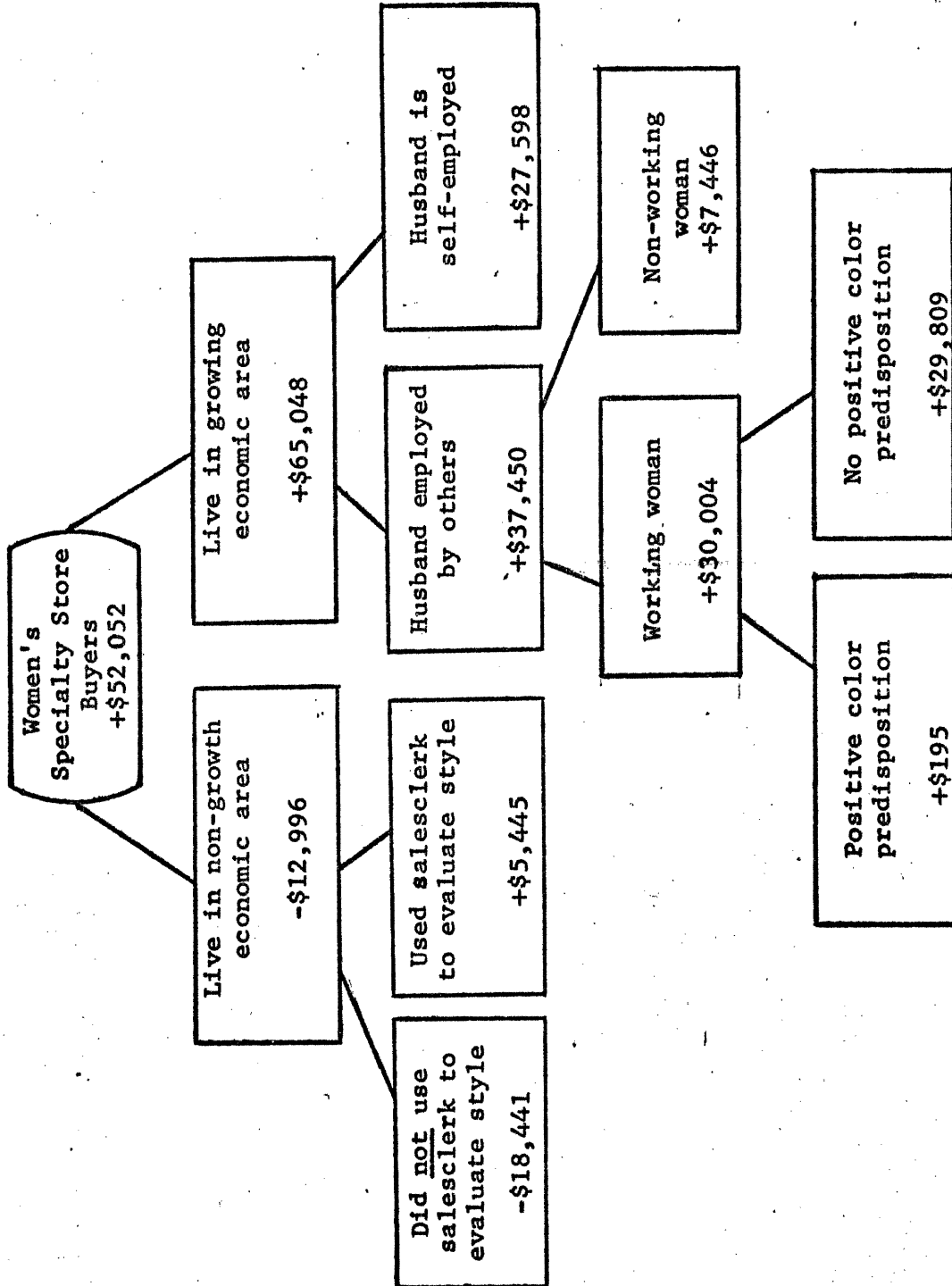
a positive color predisposition? In such cases, however, the SIMS analysis does point out the cost of that segmentation failure.

COMPARISON OF SIMS AND AID

To make a comparison of the SIMS algorithm with AID's, we took the output of the AID analyses and injected the same cost and revenue calculations to produce the dollar figures seen in Figures 6 and 7. To make the comparison equitable, we used in each case the same number of iterations for the AID trees as evolved from the SIMS analysis. The contribution amounts were assigned to each box in the AID analyses by the following method: in the tree for department store buyers (Figure 2) there is a ratio of 21/214 in the first left-hand box--"paid for garment by cash"; we multiplied the buyers by the contribution of \$6.23 and the nonbuyers ($214 - 21 = 193$) by the \$.89 cost-of-selling and then multiplied by the population ratio of 284.1882, thus arriving at the net contribution to overhead and profit (or in this case, cost) of -\$11,635.

There are two facts that emerge from comparing these modified AID trees to the SIMS analyses. First, if we were to limit selling to just those branches of each tree that terminate in a positive amount, the SIMS analyses would be more efficient because they give a higher contribution to overhead and profit. Second, in the case of the AID analyses there are many "splits" of buyers into two positive groups--not really helpful to the strategist trying to subdefine his market. By comparison, the SIMS trees always split into positive and negative groups, thus giving the marketer a more optimal segmentation of his market. To look further at the first point, let us compare the SIMS analysis (Figure 4) to the AID analysis (Figure 6) for department-store buyers: the maximization of profit and overhead contributions by selling only to the positive tree branches are \$80,936 for SIMS and \$70,818 for AID; similarly, for women's specialty store buyers, the maximum contribution from SIMS is \$80,894 and only \$70,493 for AID.





SUMMARY

We began this study by questioning whether AID's least-squares algorithm, despite its recent widespread use, forms a market segmentation that meets the profit objectives of most marketers. The use of SIMS with its simple cost-profit formulation as an alternative seems justified when the results are compared to those produced by AID. Certainly we recognize that more sophisticated cost data could be used; for example, good marketing cost analysis could reveal more clearly the costs of trying to market to nonbuyers. We offer this simplistic version as an incentive to further such development.

FIGURES:

1. Buyer construct. (page 2)
2. AID tree diagram--independent department store buyers. (page 5)
3. AID tree diagram--women's specialty store buyers. (page 6)
4. SIMS tree diagram--independent department store buyers. (page 13)
5. SIMS tree diagram--women's specialty store buyers. (page 14)
6. Modified AID tree--independent department store buyers. (page 17)
7. Modified AID tree--women's specialty store buyers. (page 18)

FOOTNOTES

1/ James N. Morgan and John A. Sonquist, "Problems in the Analysis of Survey Data and a Proposal," Journal of the American Statistical Association, 58 (June, 1963) 415-34; and John A. Sonquist, Elizabeth L. Baker, and James N. Morgan, Searching for Structure (Alias - AID-III) (Ann Arbor: Survey Research Center, Institute for Social Research, The University of Michigan).

2/ Dennis Gensch and Richard Staelin, "The Appeal of Buying Black," Journal of Marketing Research, IX (May, 1972), 141-48; Joseph Newman and Richard Staelin, "Prepurchase Information Seeking for New Cars and Major Household Appliances," Journal of Marketing Research, IX (August, 1972), 249-57; Henry Assael, "Segmenting Marketing by Group Purchasing Behavior: An Application of the Aid Technique," Journal of Marketing Research, VII (May, 1970), 153-58; William Wilkie, "Extension and Tests of Alternative Approaches to Market Segmentation," Working Paper No. 323 (Lafayette, Ind.: Institute for Research in the Behavioral, Economic, and Management Sciences, Purdue University, September, 1971).

3/ John A. Howard and Jagdish N. Sheth, The Theory of Buyer Behavior (New York: John Wiley & Sons, 1969).

4/ U.S., Bureau of the Census, Census of the Population, 1970, PC (1), A27, B27, C27 (Washington, D.C.: Government Printing Office, 1971).