both of U.S. DOT, are gratefully acknowledged for assistance in the running of computer programs and the preliminary analysis of results. We accept full responsibility for the accuracy of the findings and the conclusions derived from the findings.

REFERENCES

1. T. F. Golob. Attitudinal Models. In Urban Travel Demand Forecasting (D. Brand and M. L. Manheim, eds.), HRB, Special Rept. 143, 1973, pp. 130-145.
2. T. F. Golob and R. Dobson. Assessment of Preferences and Perceptions Toward Attributes of Transportation Alternatives. In Behavioral Demand Modeling and Valuation of Travel Time (P. R. Stopher and A. H. Meyburg, eds.), TRB, Special Rept. 149, 1974, pp. 58-81.
3. R. Dobson. Uses and Limitations of Attitudinal Modeling. In Proc., 2nd International Conference on Behavioral Travel Demand (P. R. Stopher and A. H. Meyburg, eds.), Elsevier, New York, 1976.
4. D. T. Hartgen. Attitudinal and Situational Variables Influencing Urban Mode Choice: Some Empirical Findings. Transportation, Vol. 3, 1974, pp. 377-392.
5. R. Dobson and J. F. Kehoe. Disaggregated Behavioral Views of Transportation Attributes. TRB, Transportation Research Record 527, 1974, pp. 1-15.
6. R. Dobson and M. L. Tischer. Beliefs About Buses, Car Pools, and Single-Occupant Autos: A Market Segmentation Approach. Proc., Transportation Research Forum, Vol. 17, 1976, pp. 200-209.
7. G. C. Nicolaidis. Quantification of the Comfort Variable. Transportation Research, Vol. 9, 1975, pp. 55-66.
8. B. D. Spear. A Generalized Attribute Variable for Models of Mode Choice Behavior. TRB, Transportation Research Record 592, 1976, pp. 6-11.
9. F. N. Kerlinger and E. J. Pedhazur. Multiple Regression in Behavioral Research. Holt, Rinehart, and Winston, New York, 1973.
10. J. Johnston. Econometric Methods. McGraw-Hill, New York, 1972.
11. J. E. Overall and C. J. Klett. Applied Multivariate Analysis. McGraw-Hill, New York, 1972.
12. P. R. Stopher. Goodness-of-Fit Measures for Probabilistic Travel Demand Models. Transportation, Vol. 4, 1975, pp. 67-83.
13. N. H. Nie, C. H. Hull, J. Jenkins, K. Steinbrenner, and D. H. Bent. Statistical Package for the Social Sciences. McGraw-Hill, New York, 1975.
14. W. J. Dixon, ed. Biomedical Computer Programs. Univ. of California Press, Berkeley, 1975.
15. R. K. Blashfield. Mixture Model Tests of Cluster Analysis: Accuracy of Four Agglomerative Hierarchical Methods. Psychological Bulletin, Vol. 83, May 1976, pp. 377-388.
16. M. I. Appelbaum and E. M. Cramer. Some Problems in the Nonorthogonal Analysis of Variance. Psychological Bulletin, Vol. 81, 1974, pp. 335-343.
17. J. Cohen. Multiple Regression as a General Data-Analytic System. Psychological Bulletin, Vol. 70, 1968, pp. 426-443.
18. R. Dobson. The General Linear Model Analysis of Variance: Its Relevance to Transportation Planning and Research. Socio-Economic Planning Sciences, Vol. 10, 1976, pp. 231-235.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.

**Mr. Dobson is now with Charles River Associates, Cambridge, Massachusetts.*

Development of Market Segments of Destination Choice

Peter R. Stopher, Transportation Center, Northwestern University

This paper discusses the development of individual-choice models of the destination choice of nongrocery shopping locations. Two key features of the approach are the use of perceptual data for characterizing alternative destinations and an attempt to segment the population before the model building on the basis of homogeneity of perceptions of destinations. Data were obtained about the perceptions of shoppers of several shopping locations and on their preferences for various attributes of the shopping locations. The attributes were selected as those that make up the image of a shopping location independent of the transportation system. Several techniques are discussed for segmenting the population by perception, all of which are based on analysis of the psychological distance between shopping locations. Given the special properties of psychological distances, two forms of analysis were undertaken. First, correlations were computed for the set of interpoint distances for each socioeconomic group identified in the data. High correlations indicate similarity of perceptual space, while low correlations indicate lack of similarity. Second, the group interpoint distances were used as inputs to an individual scaling process that attempts to fit the perceptions into a common perceptual space by stretching or compressing the axes of the space by obtaining weights on each axis for each observation. In this case, market segmentation was sought through a hierarchic, fusion clustering process on the axis weights for each socioeconomic group. The results of these analyses converge well. Length of residence and age were found to be important segmentation variables. Sex and income were not found to be very powerful segmentation variables, but occupation may be worth study as a basis for segmentation.

In the last decade, there have been numerous developments in the formulation, refinement, and operation of travel-demand models at the level of the individual trip maker (1), by using behavioral constructs from psychology and microeconomics. For various reasons, most of this work has taken place in the subchoice of travel mode, primarily that for the work trip. From time to time, extensions of the behavioral approach beyond the modal-choice process have been proposed, but little progress had been made in such extensions until recently.

A major problem in achieving such extensions is the characterization of the elements of utility of other subchoices. In modal-choice models, utility was characterized initially in terms of the physical attributes of alternative modes of travel, such as travel times and costs and frequency. Although these physical attributes provide only an incomplete specification of alternative travel modes, they have proved sufficient to allow much progress in the development of modal-choice models. Even here, however, recent research (2, 3, 4, 5) has used psychometric techniques to add nonphysical measures to the modal-utility specification. The use of simple physical measures, such as those in modal choice, has not generally appeared to be appropriate for extensions to other travel choices. Standard transportation planning procedures have characterized the attractiveness of destinations (aggregated into geographic zones) by the number of trips they attract or by variables such as floor area and employment, i.e., size measures only. Similarly, the decision to make a trip has been modeled as an aggregate phenomenon based on a very restricted set of variables, such as automobile ownership or population (6). Other pertinent choices, such as time of day and frequency, have not been a part of the traditional transportation planning process. Attempts to develop realistic models of these choices have been hampered by a lack of variables suitable for the characterization of the determinants of these choices (7) and have largely had to await some alternative method of investigation and the measurement of pertinent variables.

The use, in the early 1970s, of psychometric techniques to aid in the further development of modal-choice models (2, 3, 4, 5) suggested a way in which other modeling areas could also be developed. Specifically, this paper reports a preparatory step toward the development of destination-choice models that are based on individual and group perceptions of alternative destinations (8, 9). The major concern here is the idea of segmenting the population before building the model, to improve the accuracy and responsiveness of the model and to increase our understanding of the choice process. The principles of market segmentation have been expounded and used extensively in marketing studies (10), but only recently has the potential of the technique been recognized in connection with travel-demand estimation (11).

The subject here is the choice of destination for nongrocery shopping trips. A choice-based survey described elsewhere (12) was carried out at several shopping locations in the north and northwest suburbs of Chicago. There were several important reasons for using a choice-based sample. The most important was the necessity that a recent nongrocery shopping trip be in the respondent's mind when responding to the survey. Given the relative infrequency of such trips, it would be necessary to approach a very large number of households to achieve a sufficient sample. Second, the available budget made interview surveys inappropriate, while a mail-out-mail-back survey seemed unlikely to bring a high enough response rate. Finally, the objects of the research did not require a generalizable model at this stage, which removed the necessity to control the biases in the choice-based procedure selected. The survey obtained data about a recent shopping trip of each respondent, the preferences of the respondent for a number of different shopping locations and for attributes of shopping locations, the perceptions of the respondents of the attributes of a number of shopping locations, and some socioeconomic details of the respondent. The analysis was conducted on approximately 7000 responses. Some slight variations in sample size occurred due to varying nonresponse rates among the different socioeconomic variables.

METHODS FOR IDENTIFYING MARKET SEGMENTS

The principal hypothesis used in this research is that socioeconomic characteristics are a reasonable basis for grouping the population according to common perceptual spaces of attractiveness of shopping locations. In other words, it is assumed (a) that persons within a given socioeconomic group are more likely to have a homogeneous perception of such an attractiveness space than are those in diverse groups and (b) that segmentation based on cognition will be useful in travel-demand research. Socioeconomic characteristics were used because of their availability from such sources as censuses, which would increase considerably the usefulness of segmentation based on them. Segmentation was based on cognition rather than on behavior for several reasons. First, behavior segmentation requires time-series data from which dynamic behavior can be mapped or, alternatively, data determined before and after a major change in a shopping opportunity, and both are extremely expensive and difficult to obtain and beyond the scope of this project. Second, segmentation based on cognition is useful and relevant if behavior is also shown to be a function of cognition. (Another phase of this project is seeking such relationships, in which cognitive segmentation can provide incremental improvements in model structure.)

The hypotheses embedded in the use of the socioeconomic variables as market segmenters are as follows:

1. Length of residence acts as a proxy variable for learning about shopping opportunities and may indicate different levels of knowledge.
2. Income may be expected to determine sensitivity to price-related variables. Low-income persons may react more strongly to price variables than do higher income persons.
3. Age may be a partial proxy for the types of products sought when shopping and also for comparative sensitivity to service variables and variety measures.
4. Sex may be a segmenter variable on the experience sought in shopping and on comparative sensitivity to most of the range of attributes except price.
5. Occupation would be expected principally to discriminate behavior and cognition between those who are employed and those who are not (including students, the retired, and housewives). Beyond this, it may be a proxy for various life-style variables.

Given these basic assumptions, the research is aimed at determining whether the finest level of groupings obtained in the survey is necessary to characterize homogeneity. The procedure adopted was, therefore, a hierarchical combination of the smallest groupings into larger groupings that yet represent homogeneity in perception. Socioeconomic variables are not the only basis for market segmentation, personality variables may be more appropriate, although less useful to the practicing transportation planner.

The appropriate subgroupings of the population are shown in Table 1. The basis of the grouping process is to obtain a perceptual space for each subgroup and then determine the similarity of the spaces among groups. The initial analysis was carried out on the basis of one socioeconomic variable at a time, without examination of two or three-way classifications of the population. These are a part of another analysis that is not yet completed, but proving extremely expensive.

The segmentation technique is based on the use of aggregate measures for each socioeconomic group. The

Table 1. Socioeconomic groups for first-cut analysis.

Characteristic	Subgroup	Dimensionality
Sex	Female	3
	Male	4, 3
Age, years	≤16	4, 3
	16 to 21	4, 3
	22 to 29	4
	30 to 39	3
	40 to 49	3
	50 to 59	4, 3
	≥59	4, 3
Income, $	≤10 000	3
	10 001 to 15 000	2, 3
	15 001 to 20 000	3
	20 001 to 25 000	4
	25 001 to 50 000	4
	≥50 000	3, 2
Occupation	Military	—[a]
	Salesperson	4
	Teacher	3
	Professional person	4, 3
	Craftsperson	3
	Clerical worker	3, 2
	Student	4
	Housewife	4, 3, 2
	Government worker	4, 3
	Retired person	4, 3
	Other	3
Length of residence, years	≤4	4, 3
	4 to 6	4, 3
	7 to 10	4, 3
	≥10	4, 3

[a]Too few responses in this category to develop multidimensional solution information.

ideal procedure would be to obtain individual perceptual spaces and group the sample on the basis of similarity of these spaces. However, individual spaces can be determined only by the Individual Differences in Orientation Scaling (IDIOSCAL) program (14), which has an upper limit of 25 individuals. Individual weights for a common space can be determined for a maximum of 100 individuals through the Individual Scaling (INDSCAL) program (15). Clearly, with 30 initial subgroups, the IDIOSCAL program would be quite infeasible, and the use of the INDSCAL program would require repeated solutions using several unrelated random samples of 100 individuals to ensure removal of small-sample biases and idiosyncracies. Convergence of solutions also could not be ensured among the separate samples. The alternative used here was the multidimensional scaling (MDSCAL) procedure, which generates average interpoint distances in the most efficient space possible (the procedure for this is described below). These average distances were then input into the INDSCAL program to provide one of the grouping procedures examined. Thus, MDSCAL is not being used as a scaling procedure, but rather as a mechanism to determine a set of interval-scaled average interpoint distances. On this basis, the scales that are represented by the axes of the solution space become irrelevant. Also, because the usual input data to the INDSCAL program are interval data, differences among the scales are irrelevant.

To understand the problems of seeking homogeneity of perceptual spaces, some understanding is necessary of the MDSCAL procedures and the results generated by these procedures. The perceptual spaces are, in this case, to be generated as aggregate spaces for a preselected group or subgroup of the population; i.e., for each socioeconomic group identified in Table 1, one aggregate perceptual space is to be developed. The aggregate, i.e., MDSCAL, procedure involves the selection of a dimensionality that is most efficient for representing the aggregate information obtained on perceived distances between the set of stimuli (shopping centers in this case). These distances can be obtained by direct questions that request information directly about the similarity that persons perceive between alternative shopping centers vis-à-vis some prespecified metric or quality. For example, in this study, the respondents were asked to rate all possible combinations of seven regional shopping centers (Woodfield, Chicago Loop, Edens Plaza, Plaza del Lago, Golf Mill, Old Orchard, and Korvette City) on a scale of one equals completely similar to seven equals completely different in response to the question, "If all the shopping centers were equally easy to get to, how similar do you think they are to each other?" Alternatively, these distances can be derived by asking the respondents to rate each of a set of shopping centers on a number of different attributes that are postulated as making up the quality or metric to be used for judging similarity. Thus, in this study, the respondents were asked to rate the seven shopping centers on a scale of good to poor for a series of attributes such as eating facilities, layout of store, prestige of store, quality of merchandise, reasonable price, ease of returning or servicing merchandise, variety or range of merchandise, availability of credit, and availability of sale items. If a set of n stimuli are used in either of these two types of questions, then the distances between the stimuli may be represented uniquely in (n - 1)-dimensional space. For example, the survey used seven shopping-center locations for the two types of questions. Thus, the interpoint distances can be represented uniquely in six-dimensional space. Significant reduction of multidimensional spaces can be achieved only for n/3 dimensions or fewer; i.e., reductions to $[(n/3) + 1]$ dimensions can always be achieved with satisfactory results even from random data on interpoint distances. However, it is extremely difficult to find a sufficiently large number of persons with a common set of shopping centers (in the sense of all being known about) having as many as seven locations. Extensions to larger numbers of shopping locations appear infeasible.

In the method used, average distances were computed for each of the identified subgroups in the population. These distances are distances between each of the seven shopping centers in the perceived space of attractiveness to shop. The first task of the analysis is to find the most efficient dimensionality in which to express the perceptual space for the attractiveness concept without distorting the perceived distances between the shopping centers. This is the procedure that the MDSCAL program performs. In collapsing the dimensionality of the space, the procedure requires that a monotonic relation be preserved between the original interpoint distances and those in each successive reduced-dimensionality space. The requirement of monotonicity is placed on the procedure, rather than a requirement of strict linearity, because the data from which the information is derived is ordinal in nature. Thus, it would not be appropriate or correct to invest ratio properties in the base data, nor to require preservation of the sizes of the intervals between stimulus points in the space in the collapsing process. In the process of developing a perceptual space through the MDSCAL program, the orthogonal axes describing the space are located arbitrarily. Thus, there is no ready mechanism for comparing the final resulting multidimensional spaces from different socioeconomic subgroups of the population with each other, since no two spaces are necessarily located in any common way. Both rotation and translation of the axes are possible from one space to another. Figures 1, 2, and 3 show three solutions from the MDSCAL process for different subgroups of the population. It is clear from these that conclusions about homogeneity or heterogeneity of subgroups cannot be drawn, given that the axes can be rotated or translated at will from one group to the next.

To be able to segment the sample, it is necessary to find a means by which alternative spaces can be compared. Two processes appeared possible from the multidimensional-scaling work. First, the multidimensional scaling results in the production of a new set of interpoint distances for the most efficient space determined. These interpoint distances, which represent average distances for members of each subgroup in the most efficient dimensionality space, can be considered as a set of candidate values that describe each subgroup in terms of the perceptual space. Thus, one may compute either a rank or a metric correlation between the sets of distances of one group and another. Since there are seven stimuli in the space, there are 21 interpoint distances that are necessary to describe each multidimensional solution, which may be used to compute either a rank (Spearman) correlation or a linear (Pearson) correlation between them. Such a measure is computed irrespective of the rotation and translation of the representation of the multidimensional space. All that one is looking for here is a correlation of the distances between each pair of points. Because the original data from which the spaces were derived is ordinal, rather than cardinal, and the procedure for developing the multidimensional space requires only the preservation of the ordinal information, it may be more appropriate to consider a rank correlation, rather than a linear correlation. However, both types of correlations were run for these data, and comparisons were made between the results obtained. In general, rank correlations might be expected to be somewhat higher than linear correlations, and this proved to be the case. One may conclude that the Spearman correlations are generally a less sensitive test of interrelation. Indeed, the results of the parallel tests were that whenever the Pearson correlations were significant, the Spearman correlations were also significant. However, there were several cases in which the Spearman correlations were significant, but the Pearson correlations were not. On the basis of the greater sensitivity of the Pearson correlations, these are the ones used below.

The second procedure for determining the comparisons between alternative attractiveness spaces involved the use of the average interpoint distances for each subgroup as an input to the INDSCAL method of analysis. The INDSCAL model is a method for developing perceptual spaces on an individual-by-individual basis. The procedure requires, however, that all individuals be fitted to a space that has common dimensionality. Thus, for example, a target space can be preselected and a determination made of how each individual can be fitted into that space by differential scaling of the relevant axes. Alternatively, the method can be used to generate its own target space as being that one that can be most readily fitted to the entire set of observations used as input. When used on individual data, the interpoint dis-

Figure 1. Two-dimensional space for clerical workers.

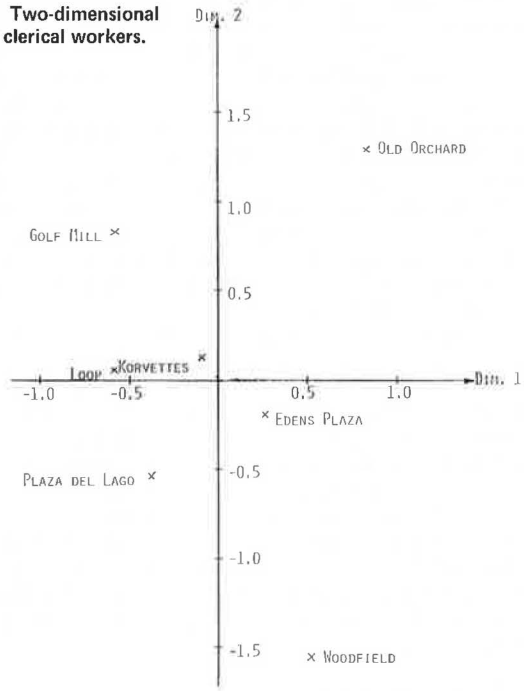


Figure 2. Two-dimensional space for incomes greater than $50 000.

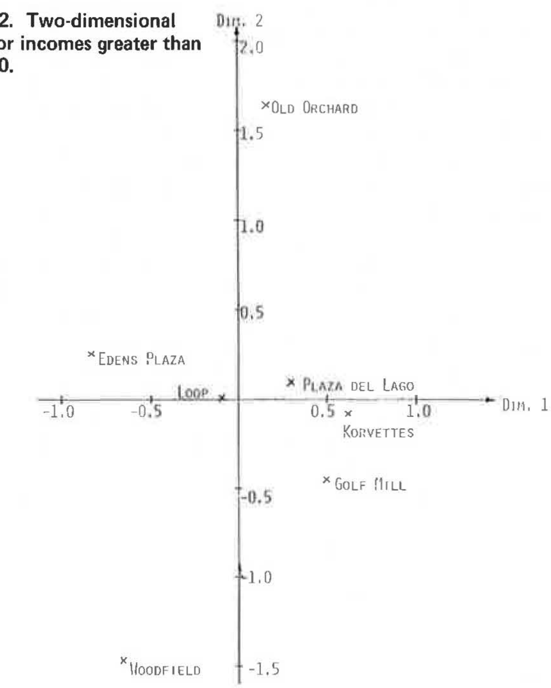
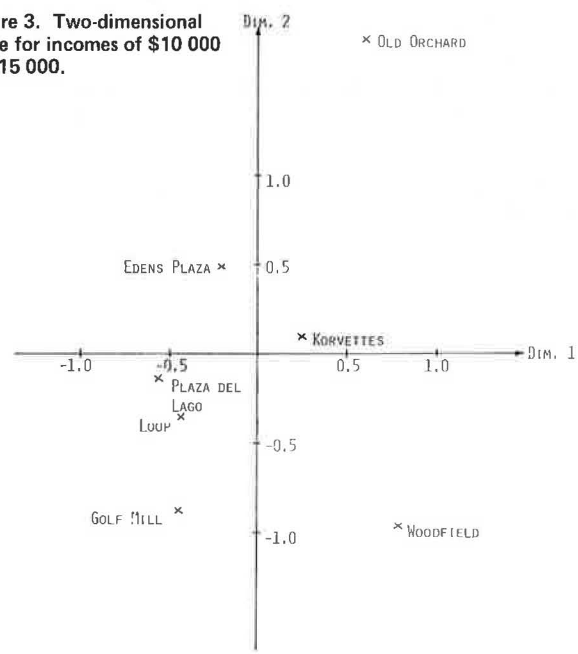


Figure 3. Two-dimensional space for incomes of $10 000 to $15 000.



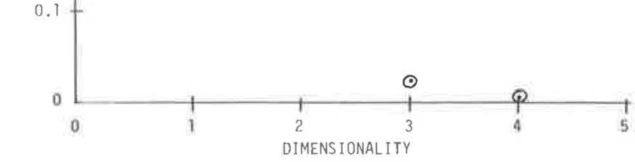
tances are computed from the original responses to the questions on either direct or derived similarities, as for MDSCAL, and a space is determined in the lowest dimensionality possible for representing all individuals. The INDSCAL procedure generates a set of weights for each individual, where these weights represent the necessary scaling of the axes for the lowest dimensionality space used, that permit each individual to be fitted into a common space with the least possible distortion of his or her own interpoint distances. The INDSCAL procedure is carried out in a metric process in which the actual distances are preserved. This description of the INDSCAL procedure is one that is relevant for its conventional use.

As outlined above, it was hypothesized that average interpoint distances derived from the MDSCAL procedure might be substituted for the individual data that would normally be the input to the INDSCAL procedure. In this manner, each of the multidimensional spaces found for the socioeconomic subgroups could be fitted to a common space, and the output weights on the various dimensions of the common space would provide a metric that could be used in some type of correlation or cluster analysis. Naturally, such a process loses the information of variance within each group, but it is not clear how serious such a loss would be here. However, there is no way in which the information can be incorporated in the process.

After each of the socioeconomic subgroups has been fitted into a common space and the weights for each of the axes of that common space have been obtained, a cluster analysis can be performed on the weights from which a hierarchy of groupings of the original subgroups can be determined. It is important, however, that neither of the methods proposed here have associated with them any statistical measures of goodness of fit.

The selection of a parsimonious space has been discussed above, but there has been no discussion of the question of how parsimony and efficiency are determined. As an aid to such selection, a statistic (stress) has been developed by Kruskal (13) that measures the degree of distortion introduced by each solution produced. Thus, as the dimensionality is reduced from the original configuration of, for example, six dimensions, a value of stress can be computed that can then be used to determine whether or not the lower dimensionality solution is acceptable. A set of empirical values has been determined for stress, in terms of specifying the degree of goodness of fit to the original data. These values are provided with descriptions in the following form: perfect fit, excellent fit, good fit, fair fit, or poor fit. Ideally, a plot of the value of stress versus dimensionality will show a characteristic elbow (Figure 4). Conceptually, this figure indicates that initially reduction in dimensionality causes no distortion in the interpoint distances, but that a point is reached at which a further reduction of one dimension causes significant distortion. It may therefore be assumed that the dimensionality immediately preceding the substantial increase in stress indicates the most efficient and parsimonious MDSCAL solution. Lower dimensionalities clearly introduce serious distortions into the data, while higher dimensionalities are not necessary, since no distortion occurs when they are reduced. The stress will not always behave in this precise fashion. It will, however, either remain approximately constant, exhibit a well-defined elbow, or have a generally upward-sloping curve as the dimensionality is reduced. In general, no other forms are possible.

Figure 4. Dimensionality versus stress for two, three, and four-dimensional solutions.



In this study, all the socioeconomic groups were run for four, three, and two-dimensional solutions. In each case, a plot was obtained of the stress versus the dimensionality, and this was used to select the appropriate dimensionality for that particular socioeconomic group. In most cases, the change of stress with dimensionality followed the ideal plot shown in Figure 4, and, the selection of the most efficient dimensionality was obvious. In some cases, however, the stress followed a more-or-less straight line that increased with decreasing dimensionality. In these cases, a solution was chosen that was based on the interpretations of fit developed by Kruskal. Where possible, the lowest dimensionality was chosen that was consistent with the empirical range for good to excellent fit. In some cases, the change in stress was such that two or more dimensionality solutions fell within the same region of fit, and in these cases, more than one dimensionality was selected as a solution. The solutions selected are shown in Table 1.

The interpoint distances from the selected multidimensional representations were then used as inputs to an INDSCAL procedure from which weights were determined for each of the subgroups. These weights were subjected to cluster analysis.

CORRELATION ANALYSIS

The first form of analysis used was the determination of the Pearson and Spearman correlations among the interpoint distances from the MDSCAL solutions. One set of correlations was determined for four-dimensional solutions, a second set for three-dimensional solutions, and a third set for two-dimensional solutions. It was not felt to be valid to compute correlations between groups whose representations were in different dimensionalities. The distinction between the two types of correlations is that the Spearman correlations are correlations only on the rank ordering of the interpoint distances, while the Pearson correlations are of a linear-regression type that are determined by assuming the distances to be metric distances. The only correlations of interest are those within a particular socioeconomic group. These are shown in Figures 5, 6, and 7. (Only the Pearson correlations are shown, because these were consistently lower than the Spearman correlations.)

In these figures, an empirical rule that may be used is that correlations below 0.5 indicate relatively little association between the variables and correlations above 0.5 indicate a fairly substantial degree of association. Thus, one may conclude that there are relatively high correlations between the sexes for the three-dimensional solutions. On the basis of the four-dimensional solutions, one could potentially place the under 16-year-old group with the 16 to 21-year-old-group, and the 16 to 21-year-old group with the 22 to 29-year-old group.

Figure 5. Pearson correlations: four-dimensional solutions.

Occupation	Sales	Prof.	Stud.	Housewife	Gov't	Ret'd
Salesman		.627	.431	.394	.308	.572
Professional			.155	.457	.143	.619
Student				.325	.432	.218
Housewife					.538	.860
Government						.288
Retired						

Income	$20-25,000	$25-50,000
$20-25,000		.643
$26-50,000		

Length of Residence	<4 yrs.	4-6 yrs.	7-10 yrs.	over 10 yrs.
<4 yrs.		.870	.566	.532
4-6 yrs.			.449	.383
7-10 yrs.				.876
over 10 yrs.				

Age Group	<16	16-21	22-29	50-59	60 and over
<16		.679	.500	.528	.574
16-21			.555	.312	.451
22-29				.546	.671
50-59					.726
60 and over					

Figure 6. Pearson correlations: three-dimensional solutions.

Sex	Male	Female
Male		.632
Female		

Occupation	Teach	Prof.	Crafts	Cler/Sec	Hswfe	Govt.	Ret'd	Other
Teacher		.522	.608	.870	.852	.409	.755	.465
Professional			.310	.422	.499	.122	.578	.433
Crafts.				.636	.537	.318	.670	.483
Cler./Sec.					.922	.496	.799	.577
Housewife						.559	.816	.390
Government							.358	.174
Retired								.487
Other								

Income	<$10K	10-15K	15-20K	>50K
<$10K		.637	.901	.598
10-15K			.586	.794
15-20K				.599
>50K				

Age	<16	16-21	30-39	40-49	50-59	60+
<16		.234	.209	.187	.306	.149
16-21			.455	.368	.422	.613
30-39				.443	.342	.644
40-49					.577	.684
50-59						.793
60+						

Length of Residence	<4 yrs.	4-6	7-10	Over 10
<4 yrs.		.723	.490	.490
4-6 yrs.			.447	.445
7-10 yrs.				1.00
Over 10 yrs.				

Figure 7. Pearson correlations: two-dimensional solutions.

Subgroup	Cler.	Hswfe	$10-15K	>$50K
Cler.		.835		
Hswfe				
$10-15K				.784
>$50K				

However, the correlation between the under 16-year-old group and the 22 to 29-year-old group is relatively low, and an optimal combination would be under 22, rather than breaking at 16. A high degree of correlation is shown between the 50 to 59-year-old group and the over 60-year-old group. Relatively high correlations seem to be demonstrated between the over 60-year-old group and all of the other age groups except the 16 to 21-year-old group. It is not completely clear why this might be so, but may indicate that this particular age category is not a useful one for discriminating perceptions of shopping opportunities. In contrast, there is a very low correlation between the under 16-year-old group and the 16 to 21-year-old group in a three-dimensional solution, and the only high correlations are those between the 40 to 49, 50 to 59, and over 60-year-old groups. In fact, the conclusion from this figure would probably be that one age group of over 40 would be sufficient to describe age groups with respect to perception of shopping-center destinations.

It does not appear to be very meaningful to consider major combinations of occupational categories. There are some quite strong correlations between certain occupational categories and very weak ones among others. For example, there are high correlations between clerical workers and teachers, between housewives and clerical workers, and between housewives and retired persons, but the conclusions that can be drawn from this are not clear. Nevertheless, the correlations are reported for completeness. The correlations based on income indicate that income is not a good discriminator of perceptions of shopping-center destinations. Indeed there are no correlations in any of these figures below 0.5, and some of the highest correlations are found in these tables. There is a very clear polarization on length of residence with a high correlation between those persons who have lived in the area less than 4 years and those who have lived in the area 4 to 6 years and a similarly high correlation between those who have lived in the area 7 to 10 years and those who have lived there more than 10 years. Both figures, which are for different dimensionalities, exhibit the same pattern. Correlations between the other pairs of groups are substantially lower, all less than 0.5. One can conclude from this that a grouping of length of residence with a break point at 6 years would appear to be appropriate. This is by far the strongest result obtained in this analysis.

CLUSTER ANALYSIS

A cluster analysis was performed on the weights for each subgroup obtained from the INDSCAL procedure by using the hierarchic fusion process (16). This analysis provided various hierarchical levels of clustering of the subgroups. Generally, only the lowest level of clustering was considered worth examining. The results of the clustering of four-dimensional solutions are shown in Table 2, and those of the three-dimensional solutions are shown in Table 3. The two-dimensional solutions were not subjected to a separate cluster analysis. On the basis of these two tables, it is again evident that length of residence may be divided at 6 to 7 years, based on the original categorization in the questionnaire. This result occurs for both the three and the four-dimensional solutions and is consistent with the results of the correlation analysis. Again, some groupings of occupations appear within the two tables, and these are generally similar to those found in the correlation analysis. For example, the correlation analysis found a high correlation between housewives and retired persons for the four-dimensional solutions, and this appears again in Table 2. Similarly, one could group clerical workers with housewives and retired persons, and the same grouping appears in Table 3. However, there is one inconsistency in the occupational groupings, in that the cluster analysis groups professional persons, craftspersons, and government workers, but these groups have very low correlations with one another.

Both the correlation analysis and the cluster analysis on INDSCAL weights showed a possible grouping, at four dimensions, of the under 16 and the 16 to 21-year-old groups. The cluster analysis did not show a grouping of those in the 50 to 59 and over 59-year-old groups. The results of the three-dimensional solutions remain consistent in grouping the under 16 and the 16 to 21-year-old groups, but this was not found so in the correlation analysis. The cluster analysis also grouped the over 59-year-old groups with the same group, a correlation that was not shown in the correlation analyses. Again, in the separate analyses, the cluster analysis shows no clustering of income groups, while the conclusion drawn from a correlation analysis was that income was a very weak determinant of perceptual differences within the population. Finally, both the correlation analysis and the cluster analysis indicate that sex is a poor discriminator of perceptual differences.

In the correlation analysis, it was not appropriate to run correlations across different dimensionality solutions. As a result, the correlation analysis has a number of gaps, where solutions are not always obtained in the same dimensionality for all subgroups. In contrast, it was reasonable to attempt a cluster analysis of the INDSCAL results combined across all dimensionalities. To do this, the INDSCAL program was run in a four-dimensional and in a three-dimensional mode, and all the MDSCAL results were input. Because the MDSCAL results used for the INDSCAL program comprise only the interpoint distances, the dimensionality of the solution does not affect the number of interpoint distances that are determined in any space. The results of the combined runs are shown in Table 4. In general, there are many consistencies across the three-dimensional and four-dimensional solutions for the combined re-

Table 2. Clustering of four-dimensional solutions within socioeconomic variables.

Original Characteristic	Cluster
Length of residence, years	
≥10	≥6
7 to 10	≥6
4 to 6	≤6
≤4	≤6
Occupation	
Salesperson	Salespersons
Professional person	Professional persons
Student	Students
Housewife	Housewives and retired persons
Retired person	Housewives and retired persons
Government worker	Government workers
Age, years	
≤16	≤22
16 to 21	≤22
22 to 29	22 to 29
50 to 59	50 to 59
≥59	≥59
Income, $	
20 001 to 25 000	20 001 to 25 000
25 001 to 50 000	25 001 to 50 000

Table 3. Clustering of three-dimensional solutions within socioeconomic variables.

Original Characteristic	Cluster
Sex	
Female	Combine sexes
Male	Combine sexes
Age, years	
≤16	≤22 and ≥59
16 to 21	≤22 and ≥59
30 to 39	30 to 39
40 to 49	40 to 49
50 to 59	50 to 59
≥59	≤22 and ≥59
Occupation	
Teacher	Teachers
Professional person	Professional persons, craftspersons, and government workers
Craftsperson	Professional persons, craftspersons, and government workers
Clerical worker	Clerical workers, housewives, and retired persons
Housewife	Clerical workers, housewives, and retired persons
Government worker	Professional persons, craftspersons, and government workers
Retired person	Clerical workers, housewives, and retired persons
Other	Other
Income, $	
≤10 000	≤10 000
10 001 to 15 000	10 001 to 15 000
15 001 to 20 000	15 001 to 20 000
≥50 000	≥50 000
Length of residence, years	
≤4	≤6
4 to 6	≤6
7 to 10	≥6
≥10	≥6

Table 4. Clustering of all solutions in three and four dimensions.

Original Characteristic	Dimensionality	Three-Dimensional Cluster	Four-Dimensional Cluster
Sex			
Female	3	Combine sexes	Combine sexes
Male	3	Combine sexes	Combine sexes
	4	Male	Male
Age, years			
≤16	3	≤16	≤16
	4	≤22	≤22
16 to 21	3	16 to 21	16 to 21
	4	≤22	16 to 21
22 to 29	4	22 to 29	22 to 29
30 to 39	3	30 to 39	30 to 39
40 to 49	3	40 to 49	40 to 49
50 to 59	3, 4	50 to 59	50 to 59
≥59	3, 4	≥59	≥59
Income, $			
≤10 000	3	≤10 000	≤10 000
10 001 to 15 000	2, 3	10 001 to 15 000	10 001 to 15 000
15 001 to 20 000	3	15 001 to 20 000	15 001 to 20 000
20 001 to 25 000	4	20 001 to 50 000	20 001 to 50 000
25 001 to 50 000	4	20 001 to 50 000	20 001 to 50 000
≥50 001	2, 3	≥50 001	≥50 001
Occupation			
Salesperson	4	Salespersons	Salespersons and professional persons
Teacher	3	Teachers, housewives, clerical workers, and retired persons	Teachers, housewives, clerical workers, and retired persons
Professional person	3, 4	Professional persons, craftspersons, and government workers	Salespersons and professional persons
Craftsperson	3	Professional persons, craftspersons, and government workers	Craftspersons and professional persons
Clerical workers	2, 3	Teachers, housewives, clerical workers, and retired persons	Teachers, housewives, clerical workers, and retired persons
Student	4	Student	Student
Housewife	2, 3, 4	Teachers, housewives, clerical workers, and retired persons	Teachers, housewives, clerical workers, and retired persons
Government worker	3, 4	Professional persons, craftspersons, and government workers	Craftspersons and government workers
Retired person	3, 4	Teachers, housewives, clerical workers, and retired persons	Teachers, housewives, clerical workers, and retired persons
Other	3	Other	Other
Length of residence, years			
≤4	3, 4	≤6	≤6
4 to 6	3, 4	≤6	≤6
7 to 10	3	≥6	≥6
	4	≥6	7 to 10
≥10	3, 4	≥6	≥10

sults, and similarly, consistency between these results and those for the separate dimensionality solutions in Tables 2 and 3. The differences between Table 4 and the results given in Tables 2 and 3 are more consistent with the results of the correlation analysis. This may be because the level of clustering is set arbitrarily in each instance, and the level at which clusters are formed and reported in Table 4 may be a higher one than that at which they are formed and reported in Tables 2 and 3. Unfortunately, there are no statistical measures that can be used to define or assess levels of clustering. One of the notable results is the clustering of incomes from $20 000 to $50 000 that is more consistent with the results of the correlation analysis. Similarly, the occupational grouping of teachers, housewives, clerical and secretarial workers, and the retired is also consistent with the results of the correlation analysis. The identification of student and other occupational categories as having no strong grouping with any other group is also borne out in both Table 4 and the earlier results of the correlation analysis. The groupings of sex, ages, and length of residence are fairly consistent between Table 4 and Tables 2 and 3 and again with the correlation analysis.

A further point of interest in Table 4 concerns the groupings of the solutions for different dimensionalities of the same attribute. In general, when the two-dimensionality solutions for the same subgroup are clustered, the selection of the lower dimensionality solution would not introduce any biases into the process; i.e., in these cases, the lower dimensionality can be considered as appropriate. This would be the case, for example, for the age groups of under 16 and 16 to 21, 50 to 59, and over 59 years. Similarly, it would be appropriate for the income group of $10 000 to $15 000 and for the occupational groups of teacher, professional person, clerical worker, housewife, government worker, and retired person. Likewise, it would be appropriate for the length-of-residence variable to be considered only at a three-dimensional solution, rather than at a four. There does not appear to be a close similarity between the three-dimensional and four-dimensional solutions for males. This suggests that a significant bias is introduced by dropping from four dimensions to three dimensions and may therefore require further analysis of whether or not sex is a good discriminating variable of perception.

CONCLUSIONS

The results of this analysis of market segmentation lead to a number of conclusions. First, both the correlation analysis and the cluster analysis of the INDSCAL weights appear to have generated convergent validities of the primary findings for grouping or not grouping among the socioeconomic characteristics. Thus, it may be concluded that the length of residence in the area and age are reasonably powerful market-segmentation variables. By and large, few new groupings of age were determined from the analysis, the only significant one being the grouping of the two lowest age groups into the single one of those individuals under 22 years old. It is possible that some of the higher age groups might potentially benefit from combination, but the two sets of analyses are not consistent on this point. It can also be concluded that sex and income do not exhibit great potential as bases for market segmentation of preferences for shopping destinations. Finally, it can be concluded that occupation may be a potential variable of segmentation, but the precise logic of its effect in defining market segments is not clear from this analysis. Subsequent work has suggested that occupation may be acting as a surrogate for other variables, such as a combined level of education and income variable, and possibly as a proxy for a stage-in-the-family-life-cycle variable, where dealing with housewives and the retired. It is also possible that some of the correlations and clustering found in the occupation variable may be spurious, due to high correlations with underlying structural variables. Thus, the original hypotheses on the socioeconomic variables have only been partially validated by this research.

Beyond this, it is clearly necessary to subject these conclusions to more stringent tests designed to determine whether or not the subgroups themselves are appropriate for market segmentation. This analysis has not addressed the question of whether these subgroups are themselves appropriate for segmenting the market, since no investigation has been undertaken of the comparative within and between-group variances, and it is not apparent how the approaches described here could be extended to covering this point.

It is also evident that testing must be undertaken on more than one-way clustering of the population. Thus, it would be appropriate to examine the possibility that two or more socioeconomic variables are needed simultaneously to define market segments in the population.

Finally, it is possible that perceptions of the attractiveness of a destination may vary with the type of goods being purchased and with an individual's knowledge of the shopping centers. Neither of these two variables were entered into the market-segmentation process reported here. It would appear appropriate to include such variables in subsequent analyses, to determine whether such variations might exist.

ACKNOWLEDGMENTS

The work reported in this paper was conducted under a contract from the Program for University Research, Office of the Secretary, U.S. Department of Transportation. The paper represents my position and not necessarily that of the U.S. Department of Transportation. The work was accomplished with much help from the research team of the project, particularly Jean M. Blin, Peter L. Watson, James Clark, and Yosy Prashker.

REFERENCES

1. P. R. Stopher and A. H. Meyburg. Behavioral Models of Travel Demand. Paper presented at the 3rd International Conference on Travel Behavior, Asheville, N.C., May 1975.
2. G. C. Nicolaidis. Quantification of the Comfort Variable. Transportation Research, Vol. 9, No. 1, 1975, pp. 55-66.
3. T. F. Golob, E. T. Canty, R. L. Gustafson, and J. E. Vitt. An Analysis of Consumer Preferences for a Public Transportation System. Transportation Research, Vol. 6, No. 1, 1972, pp. 81-102.
4. B. D. Spear. The Development of a Generalized Convenience Variable for Models of Mode Choice. Cornell Univ., PhD thesis, June 1974.
5. P. R. Stopher, B. D. Spear, and P. O. Sucher. Toward the Development of Measures of Convenience for Travel Modes. Transportation Research Record 527, 1975, pp. 16-32.
6. P. R. Stopher and A. H. Meyburg. Urban Transportation Modeling and Planning. Heath, Lexington, Mass., 1975.
7. A Disaggregated Behavioral Model of Urban Travel Demand. Charles River Associates; Federal Highway Administration, March, 1972.
8. K. P. Burnett. The Dimensions of Alternatives in Spatial Choice Processes. Geographical Analysis, Vol. 5, No. 3, 1973, pp. 181-204.
9. P. E. Green and V. R. Rao. Applied Multidimensional Scaling. Holt, Rinehart, and Winston, New York, 1972.
10. R. E. Frank, W. E. Massey, and Y. Wind. Market Segmentation. Prentice-Hall, Englewood Cliffs, N.J., 1972.
11. D. A. Hensher. Use and Application of Market Segmentation. In Behavioral Travel-Demand Models (P. R. Stopher and A. H. Meyburg, eds.), Heath, Lexington, Mass., 1976, Chap. 15.
12. P. R. Stopher and P. L. Watson. Destination Choice Modeling: An Application of Psychometric Techniques. Paper presented at the Annual Meeting of the American Psychological Association, Chicago, Ill., Sept. 1975.
13. J. B. Kruskal. Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. Psychometrika, Vol. 29, 1964, pp. 1-27.
14. J. D. Carroll and J. J. Chang. IDIOSCAL (Individual Differences in Orientation Scaling), A Generalization of INDSCAL Allowing Idiosyncratic Reference System as Well as an Analytic Approximation to INDSCAL. Paper submitted to the Spring Meeting of the Psychometric Society, Princeton, N.J., March 30 and 31, 1972.
15. J. D. Carroll and J. J. Chang. Analysis of Individual Differences in Multidimensional Scaling Via an N-Way Generalization of Eckhart-Young Decomposition. Psychometrika, Vol. 35, Sept. 1970.
16. B. Evert. Cluster Analysis. Halsted Press, Wiley, New York, 1974.

Publication of this paper sponsored by Committee on Traveler Behavior and Values.