SCALE ANALYSIS IN CONSUMER SPATIAL BEHAVIOR

AN ANALYSIS OF INTRA URBAN CONSUMER SPATIAL BEHAVIOR

AT VARYING LEVELS OF SPATIAL AGGREGATION

by

SAMUEL JOHN FULTON, B.Sc.

A Thesis

Submitted to the Faculty of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Arts

McMaster University

December 1969

MASTER OF ARTS (1969)
(Geography)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: An Analysis of Intra Urban Consumer Spatial Behavior at Varying Levels of Spatial Aggregation.

AUTHOR: Samuel John Fulton, B.Sc. (Queen's University, Belfast)

SUPERVISOR: Professor D. L. Anderson

NUMBER OF PAGES: viii, 97

SCOPE AND CONTENTS:

Intra urban consumer spatial behavior is investigated at three scales of analysis. A stepwise compilation of the regression model is used to explain variation in both distance and frequency measures at these scales. Changes in the importance of the variables between scales are noted together with changes from one area to another at the same scale of analysis. Residuals from the regression models are investigated to find the strength of locational differentials among the population in explaining consumer spatial behavior at the three scales of analysis.

ACKNOWLEDGEMENTS

This thesis could not have been completed without the assistance of a number of people. I wish to express my gratitude to my supervisor, Professor Anderson, for his useful criticism, advice and encouragement. Also I thank Dr. Peter Ambrose, and Professor Georges Potvin for many useful suggestions in the early part of the work.

I am indebted to Mr. Harold Fritz who assisted in the preparation of the figures, and to Miss Raymonde Thibeault for typing the final manuscript.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Measuring the degree of correspondence among two or more spatial patterns is a recurring theme in geographic research. Measurement may be necessary to test hypotheses regarding functional relationships between the phenomena in question, or it may be an exploratory route through which a researcher seeks to induce relationships concerning the spatial covariance of the phenomena.

The early approach to this type of analysis was map comparison and in some cases map overlays. The more powerful techniques of regression and correlation analysis are now available. These techniques were first introduced into geographic research by Kendall[1] in 1939, who although not a geographer, used a linear regression model to investigate crop productivity in England. Since this initial work, regression and correlation techniques have been further extended and are now widely used in geographic research. In most cases the major concern is with the level of explained variation in the dependent variable, as represented by the R^2 value, and the isolation of the variables

---

[1] M. G. Kendall, "The geographical distribution of crop productivity in England", Journal of the Royal Stat. Soc., 102, 1939, pp. 21-62.

contributing most to this value.

1.1 Individual and Aggregate Data

Most of the data which are available to geographers, except that gathered in the field for personal use, are aggregate data. Commonly each datum pertains to an aggregate of individuals within an areal unit, rather than to a single individual. This creates problems when statistical inferences are made. A continuum of sizes for the areal units may exist, and one of the major concerns of geography should be to systematically investigate variation at these different scales of spatial aggregation.

1.2 Scale of Analysis

It is commonly hypothesised that the relative importance of the factors found in an analysis of spatial patterns varies with the scale of analysis. Harvey[2] in a study of pattern, process and the scale problem, states that different processes become significant to the understanding of spatial patterns at different scales. A corollary to this statement is that at different scales, some processes may appear more important due to the relative invariance of others at that scale.

At present we possess little knowledge of the scale at which a particular process contributes most to the formation of a spatial pattern. It follows that when seeking to explain micro variations in pattern, the relevant processes may be different from those that we

---

[2] D. W. Harvey, "Pattern, process and the scale problem in geographical research", Trans. Inst. Brit. Geog., 42, 1968, pp. 71-78.

would consider in seeking to explain macro variations. An accurate knowledge of the scale at which a process is relatively most important, or a pattern, the most revealing, may be used to determine the scale at which a particular pattern should be measured. Hagerstrand[3] incorporates this scale finding problem into his simulation model for information diffusion. He finds that measurement should incorporate a cell size equal to the mean information field. In Hagerstrand's model this equalled a five by five matrix of his basic grid. We should likewise measure the correlation between variables at the scale of analysis at which the relationships are to be utilized or "exploited".

The problem of identifying the scale at which functional relationships or causal mechanisms are most relevant, has been considered by several researchers. McCarty, Hook and Knos[4] investigate the problem with regard to the location and distribution of certain manufacturing industries in the United States and Japan. Yule and Kendall[5] investigate the problem using agricultural data, and Warntz[6] investigates incomes. The problem can be met in any field where areal data are collect-

---

[3]T. Hagerstrand, "A Monte Carlo approach to diffusion", European Journal of Sociology, 6, 1965, pp. 43-67.

[4]H. H. McCarty, J. C. Hook and D. S. Knos, The Measurement of Association in Industrial Geography, Iowa City, 1956, University of Iowa, Department of Geography.

[5]G. U. Yule and M. G. Kendall, An Introduction to the Theory of Statistics, New York, 1950, Hafner Publishing Co.

[6]W. Warntz, Macrogeography and Income Fronts, Regional Science Research Institute, Monograph Series #3, 1965.

ed and functional relationships derived. It is unlikely that the correlation between two variables is the same when considering an area the size of the United States, and an area of county size.

1.3 Ecological and Individual Correlation

The distinction between ecological and individual correlations must be made. In an ecological correlation the variables are descriptive properties of groups, while an individual correlation is a correlation in which the variables are the descriptive properties of individuals[7]. The behavior of individuals refers to the variables describing the properties of individuals, while ecological data refer to the variables describing the properties of groups. Robinson[8], Duncan, et. al.[9], Thomas and Anderson[10] and Curry[11] have all been involved in the search for a true inference from spatially aggregated data. However, as King[12] states, no pleasing conclusion has been reached in this search.

---

[7]W. S. Robinson, "Ecological correlation and the behavior of individuals", American Sociological Review, 15, 1950, pp. 351-357.

[8]Ibid.

[9]O. D. Duncan et. al., Statistical Geography; Problems in Analysing Areal Data, New York, 1961, The Free Press.

[10]E. N. Thomas and D. L. Anderson, "Additional comments on weighting values in correlation analysis of areal data", Annals, Association of American Geographers, 55, 1965, pp. 492-505.

[11]L. Curry, "A note on spatial association", The Professional Geographer, 18, 1966, pp. 97-99.

[12]L. J. King, "The analysis of spatial form and its relation to geographic theory", Annals, Association of American Geographers, 59, 1969, p. 580.

With individual correlation the aggregation problem as investigated in the literature above does not exist. However, the problem still does exist that when looking for functional relationships, a sample of individuals has to be taken on which to base inferences pertaining to the entire population. Although a functional relationship may hold true for the population looked at in toto, it may not be true if an areal subset of the population is analysed. Thus to understand the relationships which pertain to any problem, analysis should be carried out for a range of different levels of spatial aggregation. The scale of analysis will dictate the relationships which will be most important at each level of analysis.

1.4 The Context of the Problem

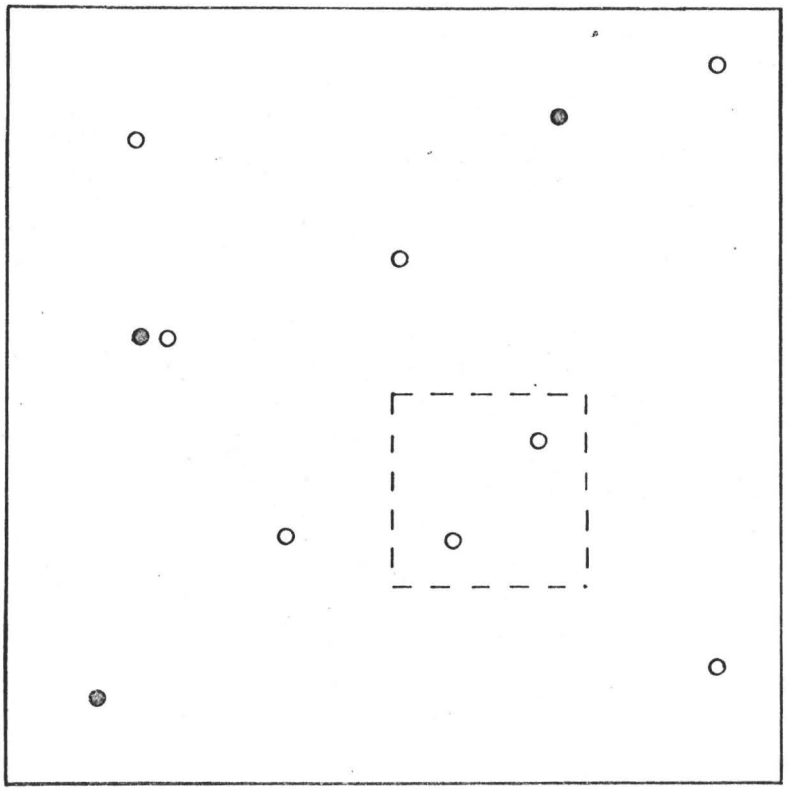
The purpose of this thesis is to investigate a spatial problem at several different areal scales. Individual consumer spatial behavior data are used. A careful interpretation of the results at the different levels is made, and differences in the functional relationships are noted in an attempt to give a comprehensive picture of consumer movement patterns as affected by the socio-economic attributes of the consumers in the sample area.

Consumer spatial behavior has been shown to be the result of socio-economic, locational and psychological differentials among the population[13]. The psychological differentials as yet have not been thoroughly examined, and little is known as to the empirical effects

---

[13] D. L. Thompson, "Future directions in retail area research", Economic Geography, 42, 1966, pp. 1-18.

FIGURE I

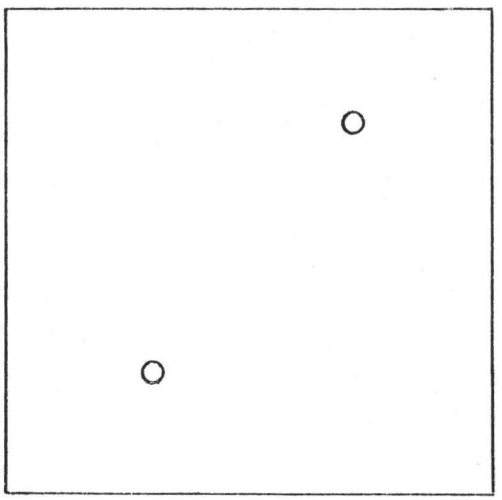


o Small family owned grocery store

● Supermarket

FIGURE II

(see inset in figure I)



they have on spatial behavior. Socio-economic and locational differentials are more easily measureable. The strength with which these factors affect spatial behavior will vary according to the scale of which investigation is carried out.

Consider the following example showing two scales at analysis. If an analysis of consumer spatial behavior is carried out for consumers living in the area shown in Figure 1, locational differentials among the population will be an important factor in discerning the outcomes of individuals' spatial behavior. This is because of the range of consumer locations which are possible in the area. However, in similarly analysing an area as shown in Figure 2, where no supermarkets exist, locational differentials will be less important because distances travelled to supermarkets are relatively uniform. Thus it is expected that socio-economic differentials may be better predictor variables in an area similar to Figure 2, whereas in Figure 1 locational variables may be better indicators.

The power of different socio-economic variables as predictors in understanding consumer movements varies from one area to another. This is because each area is made up of a unique set of attributes. This hypothesis is expanded later in the thesis. Thus the purpose of this thesis is to investigate the variation in consumer spatial behavior at several different scales, and to interpret the results in light of the inherent attributes of each area.

CHAPTER II

REVIEW OF CONSUMER BEHAVIOR LITERATURE

This review is selective in nature, drawing only on literature which is relevant to the study reported in the latter part of this thesis.

Traditionally the literature regarding consumer spatial behavior was orientated to descriptions of the size and directions of shopping movements for purposes of trade area delimitation, or the recognition of urban spheres of influence[1]. In geography, the distance travelled by consumers is used as a measure of the importance of central place functions, as represented formally by central-place theory[2].

In recent extensions of central-place theory to a broader theory to cover all tertiary activity, the more complex structure of intra city shopping movements is incorporated[3]. Broad patterns similar to those

---

[1]These have generally been stated in the form of gravity models. Review of early literature on gravity models in G.A.P. Carrothers, "An historical review of gravity and potential models", Journal Amer. Inst. of Planners, Vol. 22, 1956. Later article by D.L. Huff, "A probability analysis of shopping centre trade areas", Land Economics, Vol. 53, 1963, pp. 81-90.

[2]W. Christaller, Central Places in Southern Germany, Englewood Cliffs, 1966, Prentice Hall, Inc.

[3]W.L. Garrison, et. al., Studies of Highway Development and Geographic Change, Seattle, 1959, University of Washington Press.

in rural areas are found, but the greater variations in retail outlets, and the greater variations in purchasing power amongst an urban population lead to greater differences in the nature of city consumer movements compared to those in rural areas. Thus there have been quite a few studies analysing the many factors which may be postulated as affecting urban consumer shopping movements.

A framework by which the consumer movement problem may be viewed is given.

2.1 Consumer Analysis

In the analysis of consumer spatial behavior there are three different approaches along which research may proceed. In essence this classification constitutes the different methods of treating the dimensions of time and space.

(1) A pattern of variation over space, at a cross-section in time. Some authors who take this approach are Reilly (1931), Marble (1959), Garrison (1959), Huff (1963), Murdie (1965), Rushton (1966, 1967, 1969), Ray (1967), Brush (1968) and Davies (1969).

(2) A pattern of variation through time for one specific small area in space.

(3) The variation over space in a pattern of temporal variation. Curry (1966) and Harvey (1968).

These approaches are all legitimate goals for geographical research in the analysis of consumer spatial behavior. The careful specification of the approach to be used in any particular problem is especially important because the relative importances of the indepen-

dent variables as factors affecting consumption are likely to vary with the type of variation studied. This is why economic analysis of the consumption of individual commodities, for example, concentrates on the price elasticity of demand when studying temporal variation, but on the income elasticity of demand when studying spatial variation.

The first approach is the one which has received the most study, as it is the one for which data are the most readily available. The gravity models of Reilly and Huff are of limited use in defining the individual space preferences of consumers, since analysis is based on the aggregation of individuals. Both measures are deterministic or partially deterministic in nature resulting in a finite trade area being apportioned to a retail trade centre. There are many different routes using this brief. The following are considered in this review as they are pertinent to the study carried out in this thesis: (1) space preference studies (2) learning theories and (3) intra-urban consumer spatial behavior.

The next approach has also been applied to consumer behavior analysis but the time period has been in all cases very short, a few years at the most. Data for a longer period are not generally available and even if they are, it is doubtful if analysis of behavior over a longer period would yield any worthwhile results since attitudes and store locations change fairly rapidly.

The final approach, namely the study of variation over space in a pattern of temporal variation may perhaps be the approach along which

analysis may proceed in the future. Harvey[4] suggests the use of spectral analysis to measure the various patterns and periodicities which are found in analysing consumer spatial behavior. It is known that shopping behavior tends to be regular and periodic. Therefore it would be possible by generalizing this periodicity to define it as a function which would describe the variation in purchasing potential for a certain grocery product. This could be described by:

$$C = A_k \cos(kt - P_k)$$

C - variation in purchasing potential over time.

A_k - purchasing potential, e.g. the amount spent on each shopping expedition for a particular commodity.

k - number of waves in a basic interval of time.

t - number of shopping expeditions for a good during a week.

P_k - phase angle.

$$f(t) = \sum_{k=1} A_k \cos(kt - P_k)$$

By plotting spectral density we may thus be able to identify the wavelengths i.e. the intervals between shopping behavior over time. As yet no empirical data on consumer purchasing have been analysed using spectral analysis, but the technique is applied by Tobler[5], Colenutt[6],

---

[4]D.W. Harvey, "Pattern, process and the scale problem in geographical research", Trans. Inst. British Geographers, 39, 1968, pp. 71-78.

[5]W. Tobler, "Spectral analysis of spatial series", Proceedings, Fourth Annual Conference on Urban Planning Information Systems. Berkeley, 1966, pp. 179-185.

[6]R.J. Colenutt, "Linear diffusion in an urban setting", Geogra-

Olsson[7], and Curry[8] in dealing with problems of a similar nature. Thus as the technique is perfected it is reasonable to assume that it may be used to solve the problems which still exist in analysing temporal variation in consumer behavior.

2.2 Space Preference Studies

The premise that space preferences are unique to each individual was first stated explicitly by Isard:

> "Psychologists and sociologists, whether speaking of a gregarious instinct, or of acquired behavior patterns, or of both, have emphasised the social nature of man and his propensity to associate with groups of various sorts. One can reason that such a propensity, acquired or instinctive, is a manifestation of a positive space preference."[9]

However few empirical studies exist where the space preferences of consumers with regard to their shopping patterns have been studied. Murdie[10]

---

phical Analysis, 1, 1969, pp. 71-78.

[7] G. Olsson, "Complementary models: A study of colonization maps", Geografiska Annaler, 50, Series B, 1968, pp. 1-17.

[8] L. Curry, "Central places in the random spatial economy", Journal of Regional Science, 7, 1967, pp. 217-238.

[9] W. Isard, Location and Space Economy, New York, 1956, p. 84.

[10] R.A. Murdie, "Cultural differences in consumer behavior", Economic Geography, 41, 1965, pp. 211-233.

and Ray[11] have analysed space preferences as they are reflected by cultural differences in the population. Murdie studies the differences in the space preferences of a sample of modern Canadians, and Old Order Mennonites. He finds not surprisingly, that the Mennonites generally patronize the closest centre while the modern Canadians are not so predictable.

Ray compares the travel behavior of two groups of modern Canadians, the French and the English Canadians in Eastern Ontario. He divides his study area into four zones, an English and French core zone, and an English and French marchland zone. He does not indicate how he came about classifying these zones as the basis for analysing the means and variances of consumer behavior from one zone to another. Tests do indicate significant differences in the structure of consumer travel between the French and English core zones, and between the combined marchlands and the combined core zones, but one wonders had these different zones been bisected or broken into many smaller areas, would the same relationships have held. Ray's conclusion is that cultural differences in consumer travel behavior are important for intermediate level goods and services, but not for low order goods.

In a more exhaustive study of consumer space preferences, Rushton[12] carries out tests to examine the relationships between the

---

[11] M. Ray, "Cultural differences in consumer travel behavior in Eastern Ontario", Canadian Geographer, 11, 1967, pp. 143-156.

[12] G. Rushton, Spatial Pattern of Grocery Purchases by the Iowa Rural Population, Studies in Business and Economics, New Series, No. 9, 1966, University of Iowa.

social and economic characteristics of households, and the spatial allocation of grocery expenditures of a dispersed rural population. Rushton draws his sample from a parent population before the social or economic characteristics are assigned to them. He stratifies his population into economic and socially discrete groups, rather than on their occurrence in distinct spatial units. Thus his results are generalizations pertaining to the entire sample population. For example one of his conclusions is that:

> "... younger farmers choose significantly smaller towns as their nearest, farthest, and maximum grocery purchase towns, and they patronize a significantly larger number of towns than do younger non-farmers."[13]

His general conclusion is that grocery spatial purchase patterns conform to meaningful patterns, which can be explained (and predicted) in terms of the socio-economic characteristics of the household. The strength of these patterns however as set out by Rushton does not have any generality. His results are only true when an area the size of the Iowa study area and having the same internal distribution, is analysed.

2.3 Learning Theories

Golledge and Brown[14] formalize the market decision process in terms of a learning model. They see individuals as having a finite number of feasible alternatives at which they may satisfy their needs.

---

[13] Ibid., p. 79.

[14] R. G. Golledge and L.A. Brown, "Search, learning and the decision making process", Geografiska Annaler, 49, 1967, pp. 116-123.

Thus from a consumer's accumulated experience he will "learn" which decision results give him the greatest pay-off. He will therefore retain the rewarding responses and delete unsatisfactory ones. Search will continue until the individual's aim, be it least-cost, minimized distance or any other goal, appears to be accomplished. Once a decision is made concerning the nature of the most favorable response, search activity will be reduced and a more or less regular pattern of movement will replace it. Based on this reasoning it is legitimate to hypothesise that the length of occupancy by a consumer at a certain location may be reflected in both the distance he travels for certain goods, and the frequency with which he travels this distance. For this reason this variable is included among the independent variables to be used in this thesis.

Golledge and Brown have extracted several theories from psychology which may be useful in describing the decision making processes of a consumer. These models are as yet untested since our knowledge is not as yet sufficient to supply information regarding all the parameters for which information is required. The models are stated in general form and deal with individuals in space rather than aggregates. Any instrumentation of these theories will most likely require the grouping together of individuals into spatial units.

2.4 Intra-Urban Consumer Spatial Behavior

Marble[15] is the first researcher to quantitatively examine

---

[15] D.F. Marble, "Transport inputs at residential sites", Papers and Proc. of the Reg. Science Ass., 5, 1959, pp. 253-265.

intra-urban consumer spatial behavior. By using a series of socio-economic variables he derives three linear regression models to predict distance travelled by consumers to retail outlets, the time spent travelling, and the frequency of trip. The factors used, and found significant are size of household, income level, age and educational level of the head of the household, and the presence or absence of school children or employed persons in the household. Application of the linear regression model produces an estimating equation which is able to explain about one half of the observed variation in trip frequency. Marble, in this case, offers the first evidence to support Isard's theoretical contention that space preferences are determined by social (and psychological) factors outside of the spatial system. Marble also includes several locational variables e.g. distance to the C.B.D., and distance to the nearest low order retail centre. These latter variables, however, only account for 14% of the variation between households. Out of his empirical study, Marble proposes a theoretical model in which he sees a consumer making a trip whose relative desirability depends upon the individual's level of knowledge regarding the possible "states of nature" (i.e. outcome of each trip which could possibly be taken by a consumer).

Garrison, et. al.,[16] using the same data as Marble analyse the length and frequency of shopping trips according to the nature of

---

[16]Garrison, et. al., op. cit., Section IV.

the trip, e.g. single or multiple purpose shopping trip, and travel to small, medium or large centres. They do not however consider a socio-economic breakdown of the population in this analysis.

The sample used in both of these studies consists of only 99 households taken from a universe population of 72,000, so that many of the breakdowns are invalidated by the small sample size.

Horton[17] investigates consumer spatial behavior from another viewpoint. He points out that consumer travel behavior in one area might be expected to yield little information about the spatial distribution of shopping trip ends of a similar consumer located in another urban area, since the latter generally has a different pattern of opportunities open to him. Thus he is saying that not only are the space preferences of individuals different, but the location opportunities which are available to each consumer are different. This problem is similar to the one investigated by Rushton by means of indifference curves[18]. The measures of the locational differentials used by Horton are: contiguity effects (i.e. proximity to stores providing complementary merchandise), location of retail outlets with respect to competition, distance to higher order outlets, distances from the store to the major traffic generators (e.g. the C.B.D.), population density around

---

[17]F.E. Horton, "Locational factors as determinants of consumer attraction to retail firms", Annals, Association of American Geographers, 58, 1968, pp. 787-801.

[18]Rushton, op. cit., Chapter III.

the store location, and their income levels. Of these variables distance to nucleated centres is the most significant. Further analysis reveals that 72% of the sample consumer movements ended at "supra-local" firms which are located within nucleated centres, and which represent the larger establishments.

Horton thus concludes that relative location is the most important factor in attracting customers to larger firms providing low order goods. The smaller firms cater mainly to consumers residing within walking distance. He also finds that the frequency distribution of trips to establishments is J-shaped[19]. The division of commercial establishments into major and minor attraction outlets also facilitates an analysis of commercial trip attraction.

The difference between small and large establishments is included in the study reported in this thesis. However the division is made by the spatial behavior of the consumers. A discussion appears in Chapter III.

Clark[20] in a re-examination of the concept of range investigates the hypotheses (a) that consumers do not always travel to the nearest centre to purchase goods, and (b) that the range of goods increases as centre size increases. His object is to test these hypotheses in the

---

[19] J-shape curve is due to many establishments attracting few consumers travelling by car, while a few establishments attract many customers.

[20] W.A.V. Clark, "Consumer travel patterns and the concept of range", Annals Association of American Geographers, 58, 1968, pp. 386-396.

intra-urban context and to compare the results with a similar study carried out in the inter-urban context by Rushton, College and Clark[21]. Clark's conclusion is that the hypothesis is substantiated for grocery purchases, but not for higher order goods. He also rejects the postulate explicit in central-place theory that the range of a good remains constant irrespective of the different levels of business centres from which it is distributed. It is discovered that consumers travel greater distances to purchase given goods in higher order centres, although Clark attributes this to multi-purpose shopping and claims that order in spatial behavior would be more evident if the distribution and purchase of combinations of goods rather than single goods are analysed.

Davies has tried to isolate the effects of consumer income differences on shopping movements[22]. His method of isolation however is entirely based on his selection of two areas, which because of their locational similarities e.g. distance from the C.B.D., and social similarities, e.g. age structure and cultural background, he considers as being similar in all but income. Thus by comparing the shopping movements of a surveyed sample of consumers from the areas, he is able to discern differences in their shopping habits. He erroneously attributes these differences to income differentials between the two samples, forgetting that the space preferences of individuals are the result of many factors,

---

[21] G. Rushton, R.G. Golledge & W.A.V. Clark, "Formulation and test of a normative model for the spatial allocation of grocery expenditures by a dispersed population", Annals, Association of American Geographers, 57, 1967, pp. 389-400.

[22] R.L. Davies "Effects of consumer income differences on shopping movement behavior", Tijdschrift voor Econ. en Soc. Geografie, 60, 1969, pp. 111-121.

both locational, socio-economic and psychological. Had a regression model been formulated for the two areas, which had included variables representing all these major sets, then the true importance of the income variable could have been evaluated with respect to the other variables which are hypothesised as affecting consumer behavior. Davies' results indicate that consumers in the higher income levels generally have higher frequencies of movement for retail purchases. In this study a regression model is used to test the validity of this conclusion.

2.5 The Scale Problem

The "scale problem" has recently been referred to by several reviewers as one of the major problems in geographical research. It has been encountered several times in this literature review, but yet has only been specifically recognized by one writer, namely Horton. Briefly the problem is that models derived from data collected over relatively large areas cannot be applied to small areas or individual establishments for predictive purposes, without risking the introduction of considerable error in predicted values. The opposite of this axiom is also true. Horton side-stepped the problem by focusing his research on individual establishments.

There are several methods of tackling this problem. The first and most complex, is to break down the different periodicities which go to make up all consumer movement, and develop a structural model from this base. This has been outlined by Harvey but as yet no attempt has been made to implement it. More use of spatial auto-correlation techniques is needed before this will be possible.

The second and less sophisticated method is by investigating the same initial hypothesis at several different scales, taking in spatial units of different size for each analysis. Results will then be available for different scales, and a better picture of reality derived. Ray and Davies, perhaps most obviously have made this mistake of isolating areal units for comparison of behavior, in order to make generalizations which will be used in a context removed from the sample area. The inferences they made (1) are dependent on their definition of suitable areal units for comparison (2) lack true generality, and (3) only define relationships which correspond closely with the internal distributions of the sample units.

In this thesis an analysis of consumer spatial behavior is made using the second method. It is unlikely that any systematic variation of relationships will be isolated for the different levels of analysis, but the complexity of the different existing relationships, when the same data are analysed at different levels of spatial aggregation, is illustrated.

2.6 Choice of Variables

It is now realized that there are a vast number of variables which affect a consumer's movement patterns in order to make purchases. Some of these are scalar in nature and hence easily measureable, others are subjective in nature and may at best be only measured ordinally.

Huff[23] has provided us with the most comprehensive review of

---

[23] D.L. Huff, "A topographical model of consumer space preferences", Papers and Proc. of the Reg. Science Ass., 6, 1960, pp. 159-173.

the factors influencing consumer movement and the decision to move. Presenting his work without any empirical support, Huff claims that his model provides a target for empirical investigations. The model specifies a set of explanatory structures and processes developed on the basis of subjective choice of relevant variables. The nucleus of the model is the consumer's desire to obtain some object. The intensity of this desire is dependent on the extent of the physiological drive of the consumer, and the character of the stimulus situation. The consumer then decides how he will satisfy his desire in terms of his value system, the elements of which include geographical location, ethical and moral code, ethnic affiliation, income, personality, sex, occupation, age, education and mental synthesising abilities. The space around the consumer enters into the decision-making process through the consumer's perception of it. Other controlling factors which affect his decision are the range of goods offered, their prices and attending services, and several factors affected by mode of travel available, travel time, travel cost, and parking availability.

By taking all these elements and putting them in matrix form, all the linkages which exist between the elements are indicated. Each linkage however is accorded the same strength, a point which reduces the validity of this method of analysis considerably. Then by summing the first to the seventh order matrices the resultant gives the degree of connectivity as well as the relative percentage of total connectivity accorded to each element. From this analysis he isolates age, personality, sex, education, mental synthesising qualities, occupation and income as the most important factors in influencing consumer behavior.

These results however have little meaning in quantitative terms because of the error recognized above.

Marble[24] in his Cedar Rapids study, in which he deals with all trips made by consumers, finds the presence of one or more children in the household statistically significant at the one percent level. Using 14 variables his model explains 49% of the observed variation in the number of trips. Marble uses both a linear and a non-linear regression model to try and explain total distance travelled by his sample population. His R^2 values indicate explanation levels of only 14% and 23% respectively. Locational variables as discussed earlier prove most significant in his model. In attempting to explain variation in time spent away from home on trips, the variables, family size, educational level of the head of the household, the presence of one or more workers in the family, and the presence of one or more school children in the family prove highly significant. As the study reported in this thesis is somewhat similar to Marble's study, it includes some of the variables which Marble finds significant. Size of family, age and education of the head of the household are included as pertinent variables. Locational variables are omitted because of their low predictive value.

Rushton[25] uses similar variables and includes number of years in dwelling unit. He provides no rationale for including this variable.

---

[24] W.L. Garrison, et. al., op. cit., Chapter 9.

[25] Rushton, op. cit.

However, Golledge and Brown[26] postulate that the search for a satisfactory outlet is a function of the length of occupancy at a specific location. On this basis it is investigated in this study.

Income is also included by Rushton, and also by Davies. Rushton's general conclusion is that:

> "... the larger the size of the householder's total expenditure, the more variable his spatial expenditure patterns"[27]

while that of Davies is:

> "... income differences ... are sufficient to cause wide variations in the numbers and directions of shopping movements and preferences for store patronization"[28].

Although these two statements are rather general in nature they do indicate that income as a variable contributes to the spatial variation among consumers.

The mode of travel used by the consumer in travelling to a purchase location is more strongly related to the distance travelled than to frequency of trip. Mode of travel is most generally found as a dichotomous variable covering walking and car travel. However to be more comprehensive public transit is included in this study.

A distinction will be made between movements to local and movements to supra local establishments if these are made by the consumer. The premise on which this two fold division is made is that used by

---

[26]Golledge and Brown, op. cit.

[27]Rushton, op. cit., p. 98.

[28]Davies, op. cit., p. 121.

Horton[29] and mentioned earlier[30].

2.7 Summary

Thus from a brief outline of the relevant literature the following variables are isolated as being significant in explaining consumer spatial behavior: number of persons in the household, education of the head of the household, gross income of household, length of occupancy by household in present location, and the age of the head of the household. Also the variables, mode of travel, frequency of trip in a given time period, and the distances travelled by the consumer to fulfil certain defined needs represent the behavioral outcomes of the socio-economic background of the consumer.

---

[29] Horton, op. cit., p. 795.

[30] This distinction will be outlined further in Chapter III.

CHAPTER III

METHODOLOGY OF STUDY

3.1 The Problem

The working hypothesis is that the relative importance of the factors affecting consumer spatial behavior varies with the scale of analysis. Analysis at one scale is inadequate, as it is known that any correlation between areally distributed attributes is dependent on the size of the area from which the data are drawn. The analysis carried out here consists in finding the correlation between the chosen dependent and independent variables at different scales, and finds how changes in the correlation coefficients and the estimating linear equations are dependent on the attributes of the area from which the data are taken.

To illustrate what we may expect from this analysis an extreme example is given. The two scales at which analysis is made are the enumeration area and the census tract. All the residents in the enumeration area have the same income, but income varies over the tract. The coefficient of variation for the variable "income" in the area considered will therefore be zero. Income will not show up as an important variable in explaining the difference in travel behavior between consumers. However, if the census tract is now analysed, income will undoubtedly explain a portion of the variation between the consumers

with regard to their spatial behavior. Hence at the two different levels of analysis the explaining variables are different in importance. In this hypothetical case the differences are considerable, whereas in most empirical cases the differences will not be as marked or as regular. The principle holds that the factors affecting consumer spatial behavior vary with scale.

A test of this hypothesis is the major part of this thesis. The varying results obtained at the different levels of spatial aggregation are discussed. This study is exploratory because of limitations in time and resources. Hence analysis is carried out at only three levels of spatial aggregation.

The spatial behavior of a sample of consumers is analysed in an intra urban context. It is at this level that concentrations of people in certain areas with similar socio-economic attributes tend to be found. Hence scale of analysis tends to be more crucial, than in the inter urban or rural context, where there tends to be greater similarity or homogeneity in the population.

3.2 Sample Good

Grocery products are used in this analysis, as they represent the most frequently bought and provided good in the consumer purchase spectrum. Also, since it is possible to distinguish several different facets in a consumer's behavior with respect to his purchasing this good, the good is useful in accurately isolating the space preferences of an individual. Grocery products also tend to be highly standardized with regard to quality, so that no location may be said to offer a higher quality good than any other. Visits to

speciality food stores are omitted from the analysis. Grocery expenditures are made frequently, and the respondent who supplies the information on which the analysis is based will therefore be able to answer accurately questions concerning his purchasing movements. Hence for all these reasons, grocery purchases are the test commodity for this study.

3.3 Sample Area

No data exist which adhere to the strict sampling framework necessary to test the hypothesis stated earlier. Theoretically the area selected for this study should contain a variety of outlets which dispense grocery products. These outlets consist primarily of two types:

(1) the supermarket

(2) the small chain, family owned and variety store.

Inspection of the distribution of all the retail grocery outlets in Metropolitan Hamilton reveals that the area which offers the best opportunity mix of the above two elements is found in South Hamilton immediately above the escarpment (see Fig. 3). The area has no large sectors where there is a scarcity of retailing grocery outlets. This area also encompasses a population of varying socio-economic characteristics, as is shown by the census statistics. This criterion is useful in helping to isolate significant differences between different groups of the sample population. In general the area close to the escarpment contains the oldest dwelling units and the area in the extreme southern edge of the city contains the more recently constructed houses.

The area chosen is also unique in Hamilton in that for the most

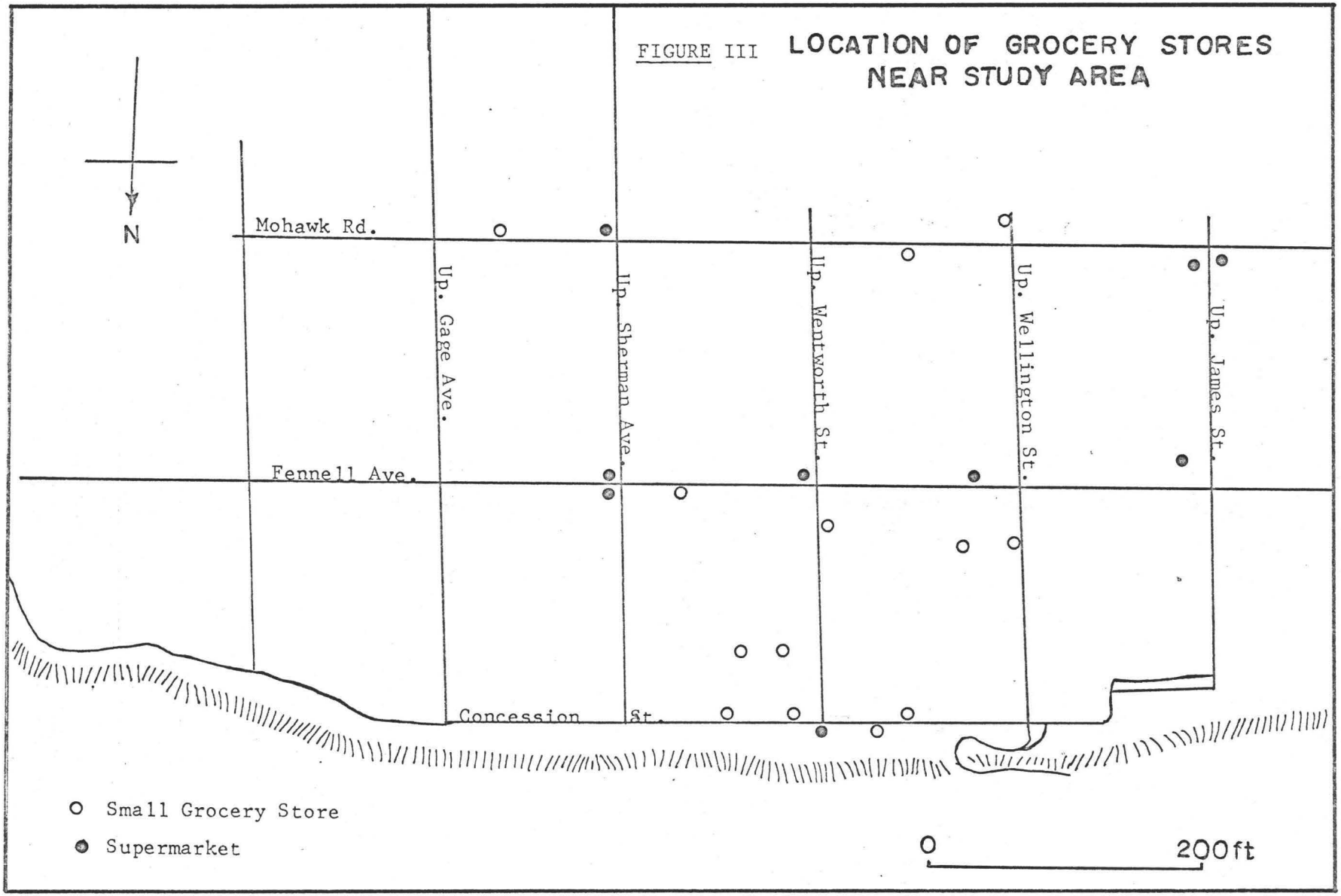


FIGURE III LOCATION OF GROCERY STORES NEAR STUDY AREA
N
Mohawk Rd.
Fennell Ave.
Up. Gage Ave.
Up. Sherman Ave.
Up. Wentworth St.
Up. Wellington St.
Up. James St.
Concession St.
Small Grocery Store
Supermarket
0
200ft

part it acts as a closed system with respect to consumer movement for grocery purchases. The majority of the population in the chosen sample area purchases its grocery needs in the area to the south of the escarpment. The "brow of the mountain"[1] acts as a strong barrier against the movement of individuals out of the sample area for such purchases.

Movements out of the area however are not ignored. An effort is made to keep them at a minimum in this study because the distance over which a consumer travels to an outlet is easier to calculate if we can assume an isotropic movement surface. Data were collected relating to both the socio-economic status and the grocery expenditure patterns of households in a small sector of an urban area.

The population is contained in census tracts 54 and 55 in the Electoral Ward of South Hamilton. A thirty percent sample was taken, according to the areal stratification given by the enumeration area boundaries. Data for the study were collected by means of a questionnaire distributed to a prearranged sample of householders.

3.4 Research Design

The analysis of consumer spatial behavior is carried out for three scales of spatial aggregation.

3.4.1 Scale I

The entire sample area consists of an area incorporating approxi-

---

[1]The "mountain" refers to the area in Metropolitan Hamilton located to the south of the limestone escarpment which runs east-west through Hamilton. The escarpment limits access to the "mountain" to only four points.

mately two square miles in a sector of South Hamilton which can be considered as a primarily residential area. Contained within this area are nine grocery stores, while 25 stores are located in the area immediately accessible to residents living in the sample area (see Fig. 3).

3.4.2 Scale II

The sample area described above is partitioned into five units. These consist of east-west divisions of the initial area. The number of stores located in these units are as follows:

Unit 1	-	3 stores
Unit 2	-	1 store
Unit 3	-	3 stores
Unit 4	-	1 store
Unit 5	-	1 store

However again it must be noted that many more stores are contained within the immediate hinterlands of these units and are not located in the actual units themselves.

3.4.3 Scale III

At this scale seventeen units are defined each consisting of approximately three to four city blocks. This is the finest scale of analysis.

3.5 The Variables

The objective in choosing relevant variables to be considered in this study is to choose those which will be the most significant in distinguishing individual's space preferences. The variables chosen can be categorized into two discrete groups, namely socio-economic variables and behavioral variables.

Pyschological variables are omitted. No study as yet has isolated empirically any such variables which are significant. It is acknowledged that the psychological variables mentioned by Huff[2] and the subjective distance variable defined by Thompson[3] affect an individual's space preferences. However these have not been incorporated in any study to date.

Locational variables such as those used by Marble[4], Horton[5] and Davies[6], although of some importance, are also omitted because of the small proportion of explanation they provide.

3.5.1 Socio-Economic Variables

The variables, after a careful review of the literature (see Chapter II), are essentially the ones which give the highest levels of explanation in former studies.

1. Number of persons in the household.

This number will reflect both the volume of grocery purchases to be made for the household, and control the amount of disposable income which is available for the purchasing of each individual's needs. It can be hypothesised to be related to both frequency of trip and distance travelled by the consumer to purchase grocery products.

---

[2]D.L. Huff, "A topographic model of consumer space preferences", Papers & Proc. of the Reg. Sci. Ass., 6, 1960, pp. 159-173.

[3]D.L. Thompson, "New concept - subjective distance", Journal of Retailing, 39, 1963, pp. 1-6.

[4]Marble, op. cit.

[5]Horton, op. cit.

[6]Davies, op. cit.

2. Wage or salary income of the household.

This variable is the amount of income received by earners in the family as cash salary and wages. It reflects the amount of income available to deal with the needs of the individual and his household.

3. Education of the head of the household.

The level of education is hypothesised to reflect the aspiration level of the consumer and hence serves as a variable for partitioning the space preferences of different individuals.

4. Length of occupancy at present address.

This refers to the period during which the head of the household has continuously occupied his present place of residence. This length is hypothesised as reflecting the amount of knowledge a consumer will possess regarding his immediate area, and of the complete urban area in which he lives. Length of occupancy in a dwelling also reflects the degree to which a consumer has established set patterns in his travel behavior[7].

5. Age of head of household.

There are several hypotheses of the relationship between age and consumer behavior. It is found as one of the most significant variables in former studies. A hypothesis tested by many authors is that mobility decreases with age.

---

[7]R.G. Golledge, L.A. Brown "Search, learning and the market decision process", Geografiska Annaler, 49, 1967, pp. 116-123.

3.5.2 Consumer Behavior Variables

Horton[8] has convincingly stated that two sets of firms are recognized by consumers. These he termed "local" and "supra-local" firms which in the context of this study refer to (1) the local isolated grocery store, and (2) the supermarket which is generally orientated towards attracting customers from a wide area. In trying to isolate components of behavior, the following measures are defined.

(1) Distance to the grocery store most frequently visited by the consumer during the last two weeks. This will give a measure of the distance a consumer will travel on those frequent low dollar trips in order to purchase a few goods to meet immediate shortages. It is reasonable to suggest that this trip is, in the majority of cases, to the nearest grocery store. Hence because of the spatial distribution of grocery outlets in the sample area there will be a division of these trips between isolated grocery outlets and supermarkets. It is more likely that this distance will correlate highly with Christaller's nearest neighbor postulate[9], since these frequent low dollar trips are in most cases single purpose trips.

(2) Are these frequent journeys (see (1) above) made by car, by bus or on foot? A direct relationship between distance and mode of travel is anticipated. At this point it is not known what percentage

---

[8]Horton, op. cit.

[9]Christaller, op. cit. p. 6.

of the sampled population will make use of public transit.

(3) Distance to the store where the largest dollar outlay is expended on a single trip per week. This will normally represent the distance to the store where the "weekly shopping" is purchased. It is most likely that this outlet is a supermarket rather than a small family store.

(4) Is this trip (see (3) above) made by car, by bus or on foot? Again as with question (2) it is expected that there will be a direct relationship between distance and mode of travel utilized, subject to certain socio-economic limitations.

(5) The frequency of all trips made to all grocery stores during the course of the last seven days. It is hypothesised that this figure is a function of the number in the household and the availability of transport.

Not all the possible hypotheses have been stated here, however during the analysis of the results obtained from the sampled population other relationships may be isolated. The significance of these relationships is evaluated for all the levels of spatial aggregation.

3.6 The Questionnaire

The questionnaire consists of a cover letter and a series of questions in two parts, (see Appendix). Part one contains questions relating to size of household, age, income, education and length of occupancy at present residential location. Part two of the questionnaire includes questions which allow the householder to identify the stores at which he makes grocery purchases, his mode of travel to these stores,

and the number of trips made in a set period of time. Rather than have a subjective estimate by each householder of the distance over which the consumer travels to his purchase location, he supplies the name of the store visited, thus standardization in distance measurement is obtained.

3.7 The Pretest

In order to test the questionnaire, and obtain an estimate of the percentage return, a sampling of thirty-five households was conducted. This pretest was undertaken in the actual study area in order to assess the speed at which interviews could be carried out in the sample area. Two blocks were selected and the thirty-five questionnaires were posted into the mailboxes of the respective households. They were collected personally three days later, as specified in the accompanying cover letter. A second contact was made the following evening to those houses where no one could be contacted on the first occasion, and to those houses where the recipients had requested that they would like the interviewer to call back. The following are the frequencies of types of responses for the pretest.

TABLE III - 1

	1st contact	2nd contact
Completed questionnaires	11	2
Refusals or incomplete	10	4
Not at home	6	4
Did not speak English	2	-
Call back requested	7	3

The final return from this pretest was 37% which is acceptable in a study of this nature. However an improvement in this figure was obtained by making several changes in the structure of the questionnaire and in the cover letter. The refusals were mostly because the information asked for was too personal. Several others mislaid the questionnaire.

Several respondents were reluctant to answer the question regarding income. However, as no surrogate for this information is very valid if the questionnaire is to be kept brief, this question could not be omitted. The order of the questions however was revised so that those pertaining to consumer spatial behavior are at the beginning, and the questions pertaining to the socio-economic status of the household follow. A "thank you" sentence was added to the end of the questionnaires.

Evening collection of the questionnaires was considered the best time for finding householders at home. This was shown by the small number of householders' in the "not at home" category in the pretest. Also, collection should take place two days at the most after the questionnaire is deposited. A longer period might encourage people to put off completion.

3.8 The Sample

The field method used in the pretest, and which has been found to be satisfactory in terms of time and effort devoted, is the random clustered sample[10]. Rather than taking a random sampling of individual

---

[10] This sampling is outlined in W.G. Cochran, Sampling Techniques. New York, 1953, pp. 234-235.

households within the study area, a cluster of household units in the form of city blocks is selected to receive the questionnaire. A time and labour consuming traverse among widely spaced individual households is thus replaced by a far less demanding method of circling a compact city block. This method also makes it easier to locate the households on a map.

The return in the main survey was higher than in the pretest. Each interviewer noted each attempted interview as either (a) a success, (b) a refusal, (c) required the interviewer to call back, or (d) the occupant did not speak English. Columns were reserved for the results of both the first and the second call back if these were necessary. If a 45% return was not obtained on first contact, return visits were made to the area in order to bring the figure up to this prescribed target[11].

Some of the questionnaires returned are incomplete and thus unusable. The final usable return consists of 343 households, representing a total of 12.5% of the total population of the sample area (see Table III-2).

3.9 Distance Measurement

Norbeck[12] has shown that if a road network between two points is fairly homogeneous and regular, it is reasonable to accept that it is enough to multiply the rectilinear distance "d" with a constant "q"

---

[11] Area here refers to each enumeration area.

[12] Stig Norbeck, "Computing distances in road nets", Papers, Regional Science Association, 12, 1963, Lund Congress, pp. 207-230.

TABLE III - 2

DISTRIBUTION OF SAMPLE HOUSEHOLDS

Location Enumeration Area	Total Pop. in Households	Sampled Households	Estimated Return	Actual Return
158	167	50	23	23
159	177	53	24	20
160	205	62	28	26
161	171	51	23	20
162	156	47	21	20
163	236.	71	32	32
164	141	42	20	18
165	145	44	21	17
166	151	45	21	21
167	139	42	19	19
168	152	46	21	17
169	153	46	21	20
170	136	41	18	15
171	161	48	21	19
172	155	47	21	20
173	137	41	18	18
174	140	42	18	18
TOTAL	2622	818	370	343

and thus obtain a good approximation of the real road distance.

$$R = q.d$$

The area from which the sample population is taken exhibits a homogeneous road network, hence the above assumption applies for all distance measurement in the area.

The extension factor "q" was calculated by locating randomly a series of points on a map of South Hamilton. These points were then paired and both the real and airline distances measured accurately. These two measures are then plotted against each other to show the relationships between the two for the area (see Fig. 4). The slope of the regression line which best fits the points is equal to the value of "q". It is estimated from the following equation:

$$q = \frac{\sum_{i=1}^{n} d.R}{\sum_{i=1}^{n} d^2}.$$

In the survey area the value obtained for "q" was 1.198, hence;

Real distance = 1.198 x Airline distance.

On this assumption the two distance measurements asked for by the questionnaire are made. The address of the householder is pinpointed on a large scale map, and the distance measured to the grocery store frequented by the respondents as set down on the questionnaires. All grocery stores were previously located on an overlay sheet. Straight line distance is taken, and this is then coded using the prearranged code as set out below. The appropriate code number is then entered on the questionnaire, to be entered later on to punched cards.

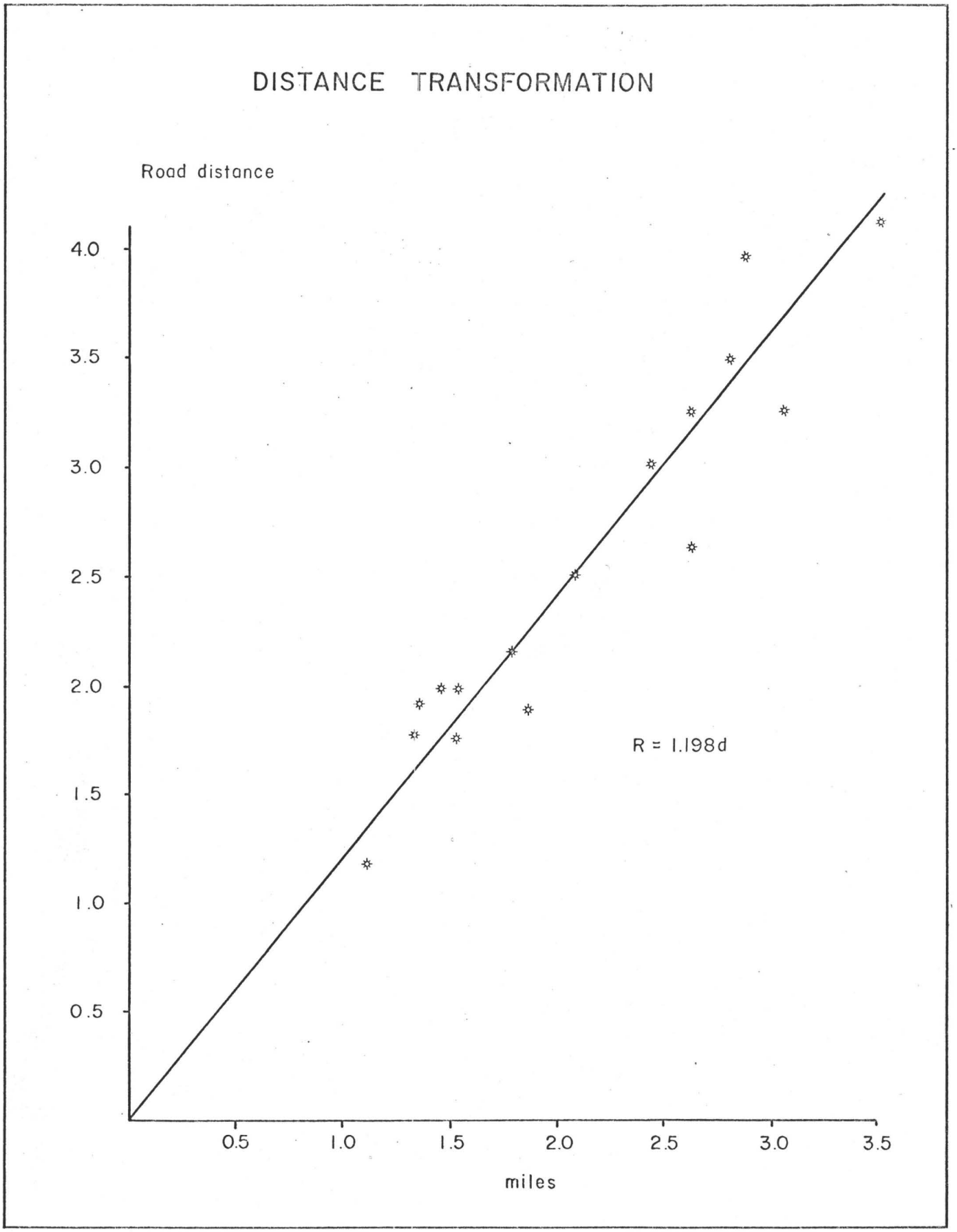

DISTANCE TRANSFORMATION
Road distance
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
R = 1.198d
0.5
1.0
1.5
2.0
2.5
3.0
3.5
miles

FIGURE IV

TABLE III - 3

Airline Distance	Real Distance	Code Entered on Data Card
0.000 - 0.333 miles	0.00 - 0.40 miles	1
0.333 - 0.666 miles	0.40 - 0.80 miles	2
0.666 - 1.000 miles	0.80 - 1.20 miles	3
1.000 - 1.333 miles	1.20 - 1.60 miles	4
1.333 - 1.666 miles	1.60 - 2.00 miles	5
1.666 - 2.000 miles	2.00 - 2.40 miles	6
Greater than 2 miles	Greater than 2.4 miles	7

CHAPTER IV

THE REGRESSION MODELS

The sample is compared with some census data. It is realized that there is a time lag between the census data (1961 and 1966) and the sample data. For the sample area the 1966 census records an average of 3.65 persons per household. The sample survey shows an average of 3.80 persons in each household. The average length of occupancy according to both the 1961 census and the sample survey is in the lower part of the six to ten year category. Enough evidence is available therefore to be confident that the sample survey is fairly representative of the sample area. The accuracy of the sample decreases somewhat when it is broken down into smaller areas. However the same degree of confidence is used for the various sub-areas.

4.1 Nature of the Variables

It is realized that the majority of the variables used in the study are not continuous and therefore cannot be truly normal. However, Hagood and Price and others[1] state that it is common practice in most of the social sciences to proceed with analysis as if distributions were

---

[1] M.J. Hagood & D.O. Price, Statistics for Sociologists, New York, 1952, Henry Holt & Co., pp. 419-433. See also, H.H. McCarty, J.C. Hook & D.S. Knos, The Measurement of Association in Industrial Geography, Iowa City, 1956, Department of Geography, University of Iowa.

normal, even when there is strong evidence to the contrary. Care must be taken however in attaching inferences to the results. Reliance can be made on relative inferences, but to a much lesser extent on absolute inferences. Normality is checked using probability paper, which has the property that the graph of a cumulative normal distribution is a straight line. It is remembered that the variables consist of class data, and as such cannot be distributed completely normal. However, this grouping is used to advantage in the isolation of inferences as explained in 4.6.2 later in this chapter.

Of the independent variables, X_7 size of household, X_8 education of the head of the household, X_9 income of the household, and X_{11} age of the head of the household are sufficiently close to the normal approximation to require no transformation. Variable X_{10}, the length of occupancy at present address, is not normally distributed in its natural state, but shows a pronounced positive skewness, which was lessened considerably by application of a reciprocal transformation. Since the "mode of travel" variables are dichotomous dummy variables (see 4.2), the problem of normality does not apply. The independent variables Y_1, distance to the most frequently visited store, and Y_2, distance to the store where the largest dollar outlay is expended on a single trip per week, are not distributed normally and no simple transformation reduces the deviations. Variable X_{14}, frequency of trip, is approximately normally distributed. Hence the only variable transformed is X_{10}, the length of occupancy at the consumer's present location.

4.2 Introduction of Dummy Variables

Regression techniques can only be applied in situations where the variables are numerically scaled. Consequently, it is necessary to redefine the mode of travel variables so that they may be inserted in the analysis on a par with the other five independent variables. To accomplish this several "dummy variables" are defined to have the following properties:[2]

M_1 - a consumer travelling on foot takes the value 1, while all other modes of travel take the value 0.

M_2 - a consumer travelling by bus takes the value 1, while all other modes of travel take the value 0.

M_3 - a consumer travelling by car takes the value 1, while all other modes of travel take the value 0.

The dummy variables M_4, M_5 and M_6 are likewise defined to deal with mode of travel on the second distance measure. Thus the regression model used is

$$Y = \alpha + M_1X_1 + M_2X_2 + \ldots + M_6X_6 + \beta_7X_7 + \ldots + \beta_nX_n.$$

The model reflects shifts of the intercept of the regressions among the three modes of travel. A fortran program was compiled to make the necessary transformations.

4.3 Summary of Variables

The following is a summary of the variables used in the analysis.

---

[2] D.B. Suits, "Use of dummy variables in regression equations", Journal of the Amer. Stat. Ass., 52, 1957, pp. 548-551.

4.3.1 Independent Variables

X_1 - Does the consumer travel the distance to the store most frequently visited, on foot (1 = yes and 0 = no)?

X_2 - Does the consumer travel the distance to the store most frequently visited, using public transit (1 = yes and 0 = no)?

X_3 - Does the consumer travel the distance to the store most frequently visited, by car (1 = yes and 0 = no)?

X_4 - Does the consumer travel on foot to the store where he spends the largest dollar outlay on a single trip during a week (1 = yes and 0 = no)?

X_5 - Does the consumer travel by public transit to the store where he spends the largest dollar outlay on a single trip during a week (1 = yes and 0 = no)?

X_6 - Does the consumer travel by car to the store where he spends the largest dollar outlay on a single trip during a week (1 = yes and 0 = no)?

X_7 - Number of persons in the household.

X_8 - Education of the head of the household.

X_9 - Income of the household as salary and wages.

X_{10} - Length of occupancy by householder at present address. (Reciprocal transformation)

X_{11} - Age of the head of the household.

4.3.2 Dependent Variables

Y_1 - Distance to the grocery store visited most frequently in the last two weeks.

Y_2 - Distance to the grocery store where the largest dollar outlay is spent on a single trip during the last week.

Y_3 - Number of trips made to all grocery stores in the last week.

4.4 Composition of the Sample

The entire sample is initially analysed to show both the composition of the area, and the relationships which exist within it. A summary of the sample is given in Figures 5 to 7.

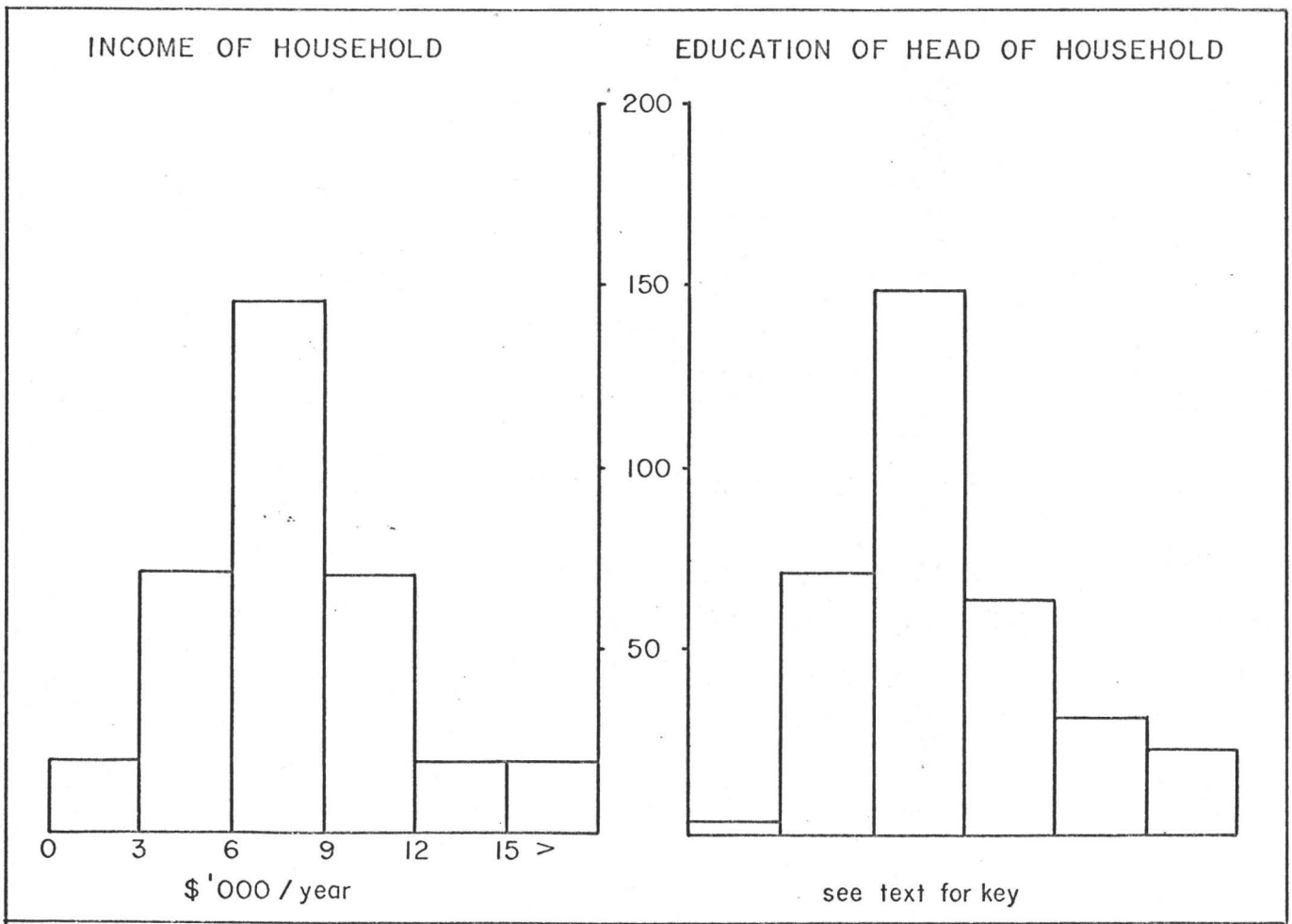
With regard to the analysis of the mode of travel, it is interesting to note that public transit is used a small number of times by respondents making journeys for grocery purchases. This may be partially explained by the fact that although public transit bus services skirt the area, they do not traverse it. Hence bus services are of limited use to the consumers in the sample area.

The means of the two distance measures substantiate the hypothesis set up in Chapter III. The distance to the most frequently visited store is 0.91 miles, while the distance to the store where the largest amount of money is spent on a single trip is 1.11 miles. The latter distance is 22% longer than the former. However 265 households or 74% of the sample travel the same distance on both occasions. This is based on both distances being contained in the same distance category, and does not necessarily imply that the consumer visited the same store on each occasion.

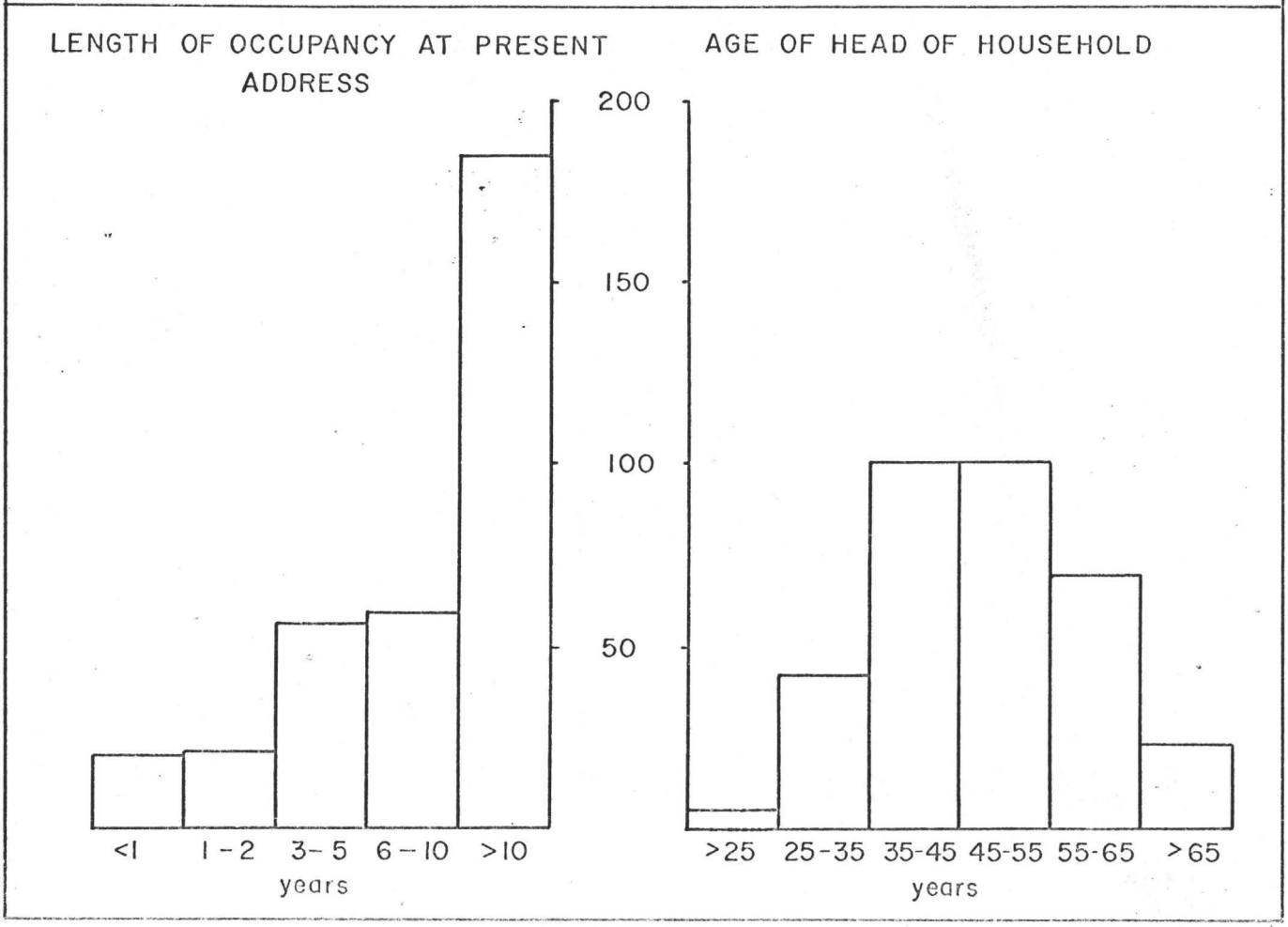
In one week the frequencies of trips for grocery purchases decline monotonically from one trip through six trips or more, (see Figure 6). This histogram is similar to the Garrison et. al.[3] findings

---

[3] Garrison, et. al., op. cit., p. 214.


INCOME OF HOUSEHOLD
200
150
100
50
0
3
6
9
12
15 >
$ '000 / year

EDUCATION OF HEAD OF HOUSEHOLD
see text for key


LENGTH OF OCCUPANCY AT PRESENT ADDRESS
200
150
100
50
<1
1 - 2
3 - 5
6 - 10
>10
years

AGE OF HEAD OF HOUSEHOLD
>25
25-35
35-45
45-55
55-65
>65
years

FIGURE V

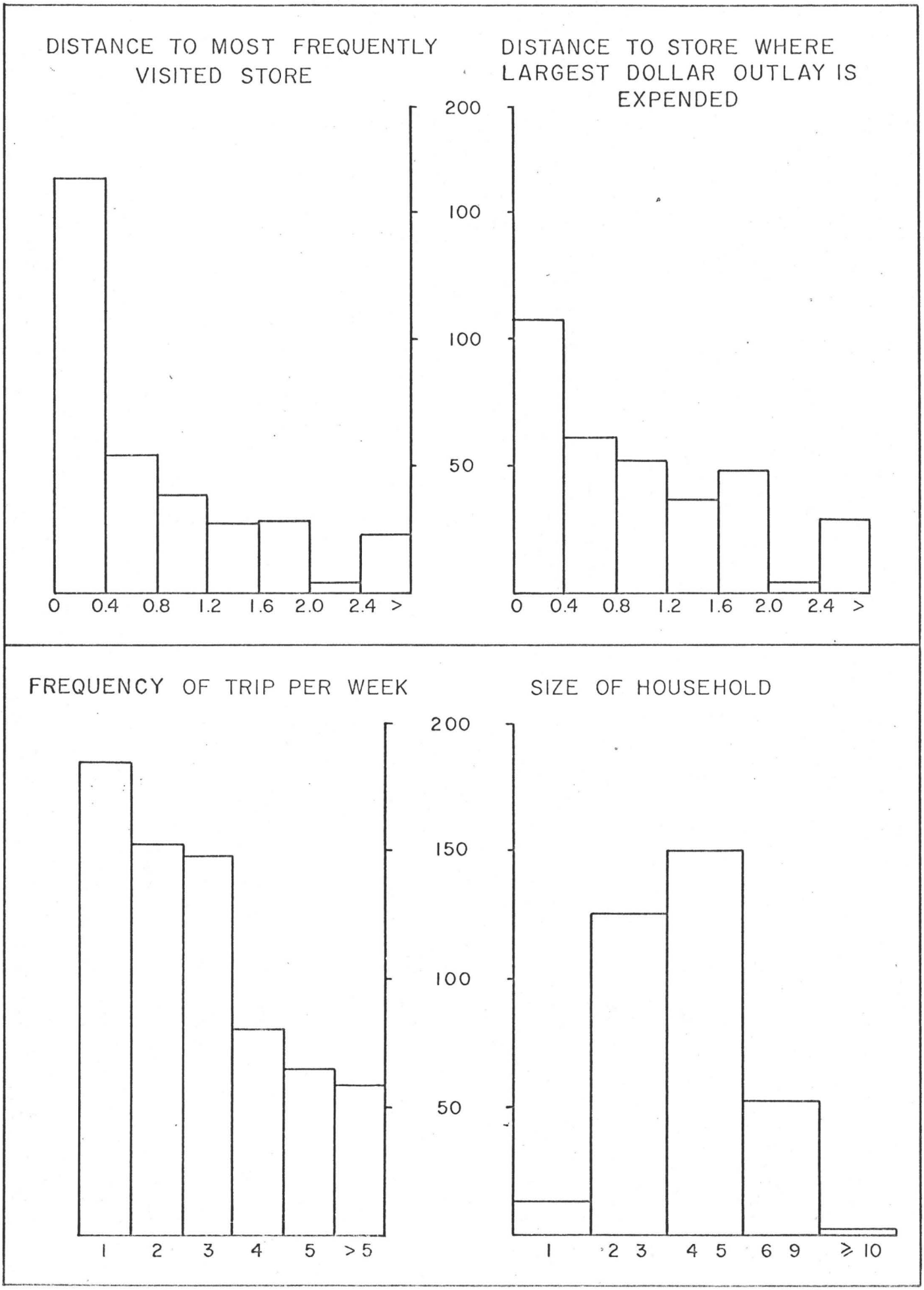
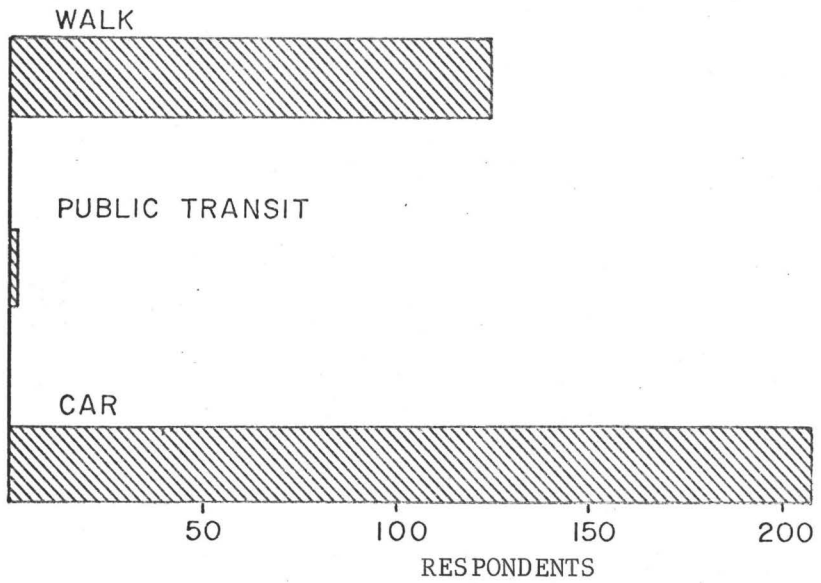

DISTANCE TO MOST FREQUENTLY VISITED STORE
200
100
50
0 0.4 0.8 1.2 1.6 2.0 2.4 >
DISTANCE TO STORE WHERE LARGEST DOLLAR OUTLAY IS EXPENDED
100
50
0 0.4 0.8 1.2 1.6 2.0 2.4 >
FREQUENCY OF TRIP PER WEEK
200
150
100
50
1 2 3 4 5 >5
SIZE OF HOUSEHOLD
1 2 3 4 5 6 9 ≥ 10

FIGURE VI

MODE OF TRAVEL

MODE OF TRAVEL TO STORE VISITED MOST FREQUENTLY



MODE OF TRAVEL TO GROCERY STORE WHERE LARGEST OUTLAY IS EXPENDED PER WEEK

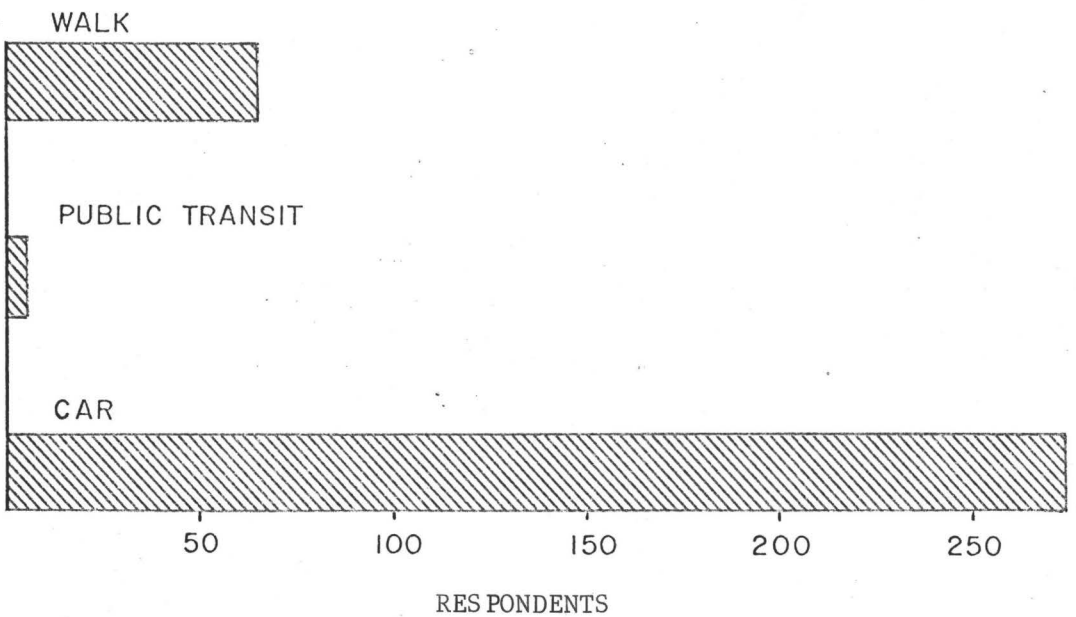


FIGURE VII

for single and multiple purpose shopping trips.

The means and standard deviations of the variables for the entire sample are summarized in Table IV-1.

TABLE IV-1

MEANS AND STANDARD DEVIATIONS FOR THE ENTIRE SAMPLE

Variable	Mean	Standard Deviation
X_1	0.3586	0.4803
X_2	0.0029	0.0540
X_3	0.6326	0.4828
X_4	0.1836	0.3879
X_5	0.0146	0.1200
X_6	0.7988	0.4015
X_7	2.7114	0.7818
X_8	3.3353	1.1451
X_9	3.1545	1.1658
X_{10}	4.0670	1.2324
X_{11}	3.7434	1.1715
Y_1	2.2770	1.6655
Y_2	2.7784	1.7581
Y_3	2.7842	1.5876

Reference should be made to the questionnaire (see Appendix) for interpretation of the means listed above.

An examination of the zero-order correlations (see Table IV-2) discloses a strong relationship between distance travelled to the most frequently visited store and the mode of travel utilized. For the distance to the store where the largest dollar outlay is expended on a single trip during a week, significant variables are mode of travel and length of occupancy by the consumer at his present location. Variables significant in explaining frequency of visits to all grocery stores are mode of travel and the number of persons in the household.

It is immediately obvious that the variation among consumers, as shown by their spatial movement patterns, cannot be explained satisfactorily in terms of a single independent variable. Hence multiple regression is used to generate a linear model which, if necessary, will include the contribution of all the independent variables to variation in the dependent variable. The computer manipulation of the regression model using a stepwise solution[4] is used, as it gives, in descending order, the variables which contribute most to variation in the dependent variable[5]. One variable is added to the model at a time, beginning with the variable exhibiting the highest zero-order correlation value. Then the partial correlation between the dependent and all the independent variables are computed. At the next step the variable having the highest partial correlation coefficient is added to the computation.

---

[4]A. Ralston & H.S. Wilf, Mathematical Models for Digital Computers, New York, 1960, Chapter 17.

[5]The program used exists on the IBM Computer System at McMaster University under the name "MLTREG".

TABLE IV-2 *

ZERO ORDER CORRELATION MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1.00	-0.04	-0.98*	0.58*	0.11	-0.59*	-0.03	-0.15	-0.20	0.01	0.15	-0.53*	-0.30*	0.33
2		1.00	-0.07	-0.03	0.44*	-0.10	-0.05	-0.06	-0.10	-0.00	0.06	-0.0	-0.02	-0.03
3			1.00	-0.57*	-0.16	0.61	0.05	0.18	0.23*	0.00	-0.14	0.53*	0.31*	-0.32*
4				1.00	-0.06	-0.94*	-0.12	-0.17	-0.26*	-0.00	0.18	-0.30*	-0.41*	0.10
5					1.00	-0.24*	0.04	0.03	-0.10	-0.03	0.05	-0.08	0.04	0.05
6						1.00	0.13	0.17	0.30*	0.03	-0.17	0.32*	0.40*	0.10
7							1.00	0.03	0.28*	-0.06	-0.33*	0.03	0.06	0.21*
8								1.00	0.49*	-0.00	-0.11	0.10	0.12	0.04
9									1.00	0.07	-0.17	0.15	0.17	0.05
10										1.00	0.45*	-0.14	-0.22*	-0.04
11											1.00	-0.15	-0.18	-0.03
12												1.00	0.71*	-0.28*
13													1.00	-0.08
14														1.00

*Correlations significant at the 5% level.

The stepwise procedure continues until all the specified independent variables are included, or the computation may be terminated when addition of further variables prove to add no additional explanation to the estimating equation.

Fitting the general linear regression model to the entire sample population, the resulting multiple regression coefficient obtained for the distance to the most frequently visited store was 0.558 which represents an R^2 value of 0.31. The equation obtained was:

$$Y_1 = 2.27 + 1.11X_3 + 1.89X_{10} + 1.76X_9 + 1.74X_2 + 1.80X_8 + 1.91X_7 + 2.04X_{11} + 2.20X_4 + 1.79X_6 + 1.00X_5 + 1.00X_1.$$

At the 5% level of significance the equation is terminated to:

$$Y_1 = 2.27 + 1.11X_3 + 1.89X_{10}.$$

The two variables represented in this equation are the use of car as a mode of travel used by the consumer to reach this store (X_3), and the length of occupancy by the householders in their present location (X_{10}). The next most significant variable is the income of the household (X_9).

Fitting a similar model to the second dependent variable, namely, the distance to the store where the largest dollar outlay is expended on a single trip per week, the resulting multiple correlation coefficient is 0.504 which gives an R^2 of 0.25. The equation[6] obtained is:

---

[6] β values are very small for the remaining variables.

$$Y_2 = 2.77 + 3.12X_4 + 4.40X_{10} + 4.07X_3 + 2.31X_1 + 2.00X_5 - 0.01X_6.$$

At the 5% level of significance the equation is reduced to:

$$Y_2 = 2.77 + 3.21X_4 + 4.40X_{10}.$$

The two variables represented in this case are the use of foot travel to travel the distance to this store (X_4) and the length of occupancy by the consumer at his present address (X_{10}).

Only 18% of the variation in the frequency of trips made to grocery stores during one week is explained. This figure is low compared with the 49% explanation found by Garrison. However the variables found significant by Garrison are of a dichotomous nature and are not included in this study. The equation obtained in this study is:

$$Y_3 = 2.78 + 2.38X_1 + 1.16X_7 + 0.75X_6 + 0.44X_8 + 0.65X_{10} + 0.53X_{11} + 0.52X_2 + 0.51X_9 + 0.51X_5.$$

The most efficient part of this estimating equation is

$$Y_3 = 2.78 + 2.38X_1 + 1.67X_7.$$

The two variables represented here are travel on foot to the store most frequently visited (X_1) and the number of persons in the family (X_7).

The conclusion of this analysis is that the variables chosen are not particularly good indicators in an area where there is such locational diversity in the population. The levels of explanation however are not particularly lower than those obtained by other researchers using similar models, except in the case of Y_3 where the explana-

tion is only 18%.

4.5 Second Scale of Analysis

For the second analysis the sample area is divided into five areas of almost similar size (see Fig. 8). It is logical to assume that the attributes of these areas are not similar to the attributes of the entire sample area in terms of means and variances. Hence the relationships which pertain to each of these areas differ from those found in the initial analysis of the entire sample area. The purpose of this analysis is to analyse dissimilarities in the factors which are the most significant in explaining the variation between consumers in each of the five areas. These are of interest to any entrepreneur, e.g. a small grocery store entrepreneur, wishing to exploit a subset of the entire sample area.

The means and standard deviations of variables in these five areas are summarized in Table IV-3. It is obvious that dissimilarities exist among the five areas. For example the mean number of persons walking to the store they most frequently visit varies from 0.13 in Area 1, to 0.53 in Area 5. This can be accounted for by the fact that consumers in Area 5 are closer to a larger number of stores (those along Concession Street - see Fig. 3) than those in Area 1, hence in the latter case a car is used more frequently to visit the store[7]. This trend is also reflected in variables X_4 and X_6. Among the socio-economic variables, variations in the mean age of the head of the household reflects

---

[7]Compare the values of X_3 for Area 1 and Area 5 to see the complement of this situation.

SCALE OF ANALYSIS

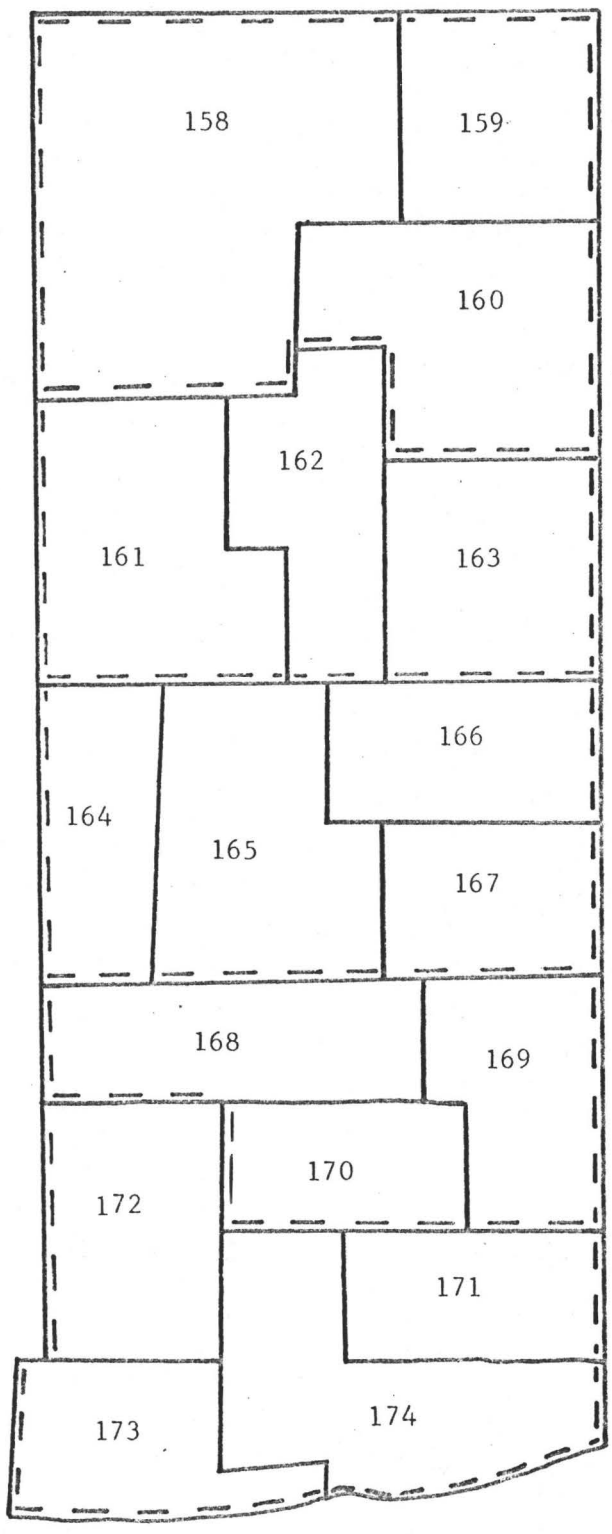


FIGURE VIII

a range from 3.41 (approx. 45 years old) in Area 1, to 4.14 (approx. 52 years old) in Area 5.

In Area 1 where auto travel is found to be used the most frequently, the distances travelled are the longest.

A stepwise multiple regression program was run on data from the five areas, each area requiring three separate runs of the program, one for each of the dependent variables used in the analysis (see Table IV-4).

4.5.1 Distance Measure I (Y_1)

The multiple regression coefficients on Y_1 for the five areas range from 0.527 to 0.698. In Area 5 which is the most northerly areal unit, almost half of the variation in Y_1 is explained by the independent variables. The lowest explanation is obtained in Area 3 where explanation is 28%. In the next chapter these differences are investigated in an attempt to explain their occurrences in terms of the composition of the population of the areas. The estimating equations which include only significant variables are as follows:

Area 1 $\quad Y_1 = 2.84 + 3.07X_1 + 4.61X_{10}$

Area 2 $\quad Y_1 = 2.08 + 1.00X_3 + 2.22X_7$

Area 3 $\quad Y_1 = 2.07 + 2.55X_1 + 1.13X_7$

Area 4 $\quad Y_1 = 2.22 + 1.17X_3$

Area 5 $\quad Y_1 = 2.19 + 1.13X_3 + 2.20X_{10}$

TABLE IV-3

VARIABLE	AREA 1		AREA 2		AREA 3		AREA 4		AREA 5	
	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.
X_1	0.13	0.34	0.40	0.49	0.32	0.47	0.44	0.50	0.53	0.50
X_2	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.12	0.00	0.00
X_3	0.87	0.34	0.60	0.49	0.67	0.47	0.55	0.50	0.46	0.50
X_4	0.03	0.17	0.20	0.40	0.12	0.32	0.26	0.44	0.34	0.48
X_5	0.00	0.00	0.03	0.16	0.03	0.17	0.02	0.12	0.00	0.00
X_6	0.97	0.17	0.77	0.42	0.85	0.35	0.73	0.45	0.64	0.48
X_7	2.72	0.73	2.79	0.74	2.83	0.73	2.61	0.78	2.58	0.99
X_8	3.42	0.99	3.39	1.06	3.30	1.08	3.18	1.20	3.37	1.43
X_9	3.33	1.04	3.03	0.91	3.22	1.08	3.08	1.18	3.14	1.61
X_{10}	4.04	1.13	3.86	1.32	4.38	1.02	4.12	1.26	3.95	1.38
X_{11}	3.41	0.97	3.44	1.15	4.00	0.95	3.85	1.14	4.14	1.47
Y_1	2.84	1.53	2.07	1.62	2.07	1.62	2.23	1.66	2.19	1.83
Y_2	3.20	1.54	2.71	1.76	2.57	1.75	2.80	1.76	2.58	1.96
Y_3	2.64	1.62	2.81	1.59	2.89	1.60	2.56	1.46	3.03	1.67

NOTE: Reference should be made to the questionnaire (see Appendix) for interpretation of the means and standard deviations listed above.

TABLE IV-4

MULTIPLE REGRESSION FOR 5 AREAS

AREA	Multiple 'r' for Distance I	Multiple 'r' for Distance II	Multiple 'r' for Frequency
1. Enumeration Areas 158/159/160	0.544 (30%)	0.419 (18%)	0.317 (10%)
2. Enumeration Areas 161/162/163	0.597 (36%)	0.570 (32%)	0.602 (36%)
3. Enumeration Areas 164/165/166/167	0.527 (28%)	0.541 (29%)	0.470 (22%)
4. Enumeration Areas 168/169/170/172	0.619 (38%)	0.610 (37%)	0.429 (18%)
6. Enumeration Areas 171/173/174	0.698 (48%)	0.635 (40%)	0.564 (32%)

* Figures in parentheses give the percentage of variability accounted for.

In all cases the dichotomous variables relating to the mode of travel chosen for the trip contribute most to explanation. Next in importance are the size of household (X_7) and the length of occupancy by the consumer at his present address (X_{10}).

4.5.2 Distance Measure II (Y_2)

The range obtained for the multiple regression coefficients with respect to Y_2 varies from a low of 0.419 for Area 1 to a high of 0.635 for Area 5 (see Table IV-4). It is legitimate to surmise that the same set of factors which explains the first distance measure explains the second distance measure. This is recognized from the estimating equations in which only those variables which are most significant are included:

Area 1 $\quad Y_2 = 3.20 + 4.58X_{10}$

Area 2 $\quad Y_2 = 4.50 + 5.18X_4 + 7.35X_7 - 9.88X_{11}$

Area 3 $\quad Y_2 = 3.86 + 7.89X_{10} + 5.33X_7$

Area 4 $\quad Y_2 = 2.80 + 1.27X_6 + 2.37X_{10}$

Area 5 $\quad Y_2 = 5.62 + 2.70X_6 + 5.44X_{10}$.

The variables X_4 and X_6, relating to the mode of travel taken by the consumer to the grocery store where he spends his largest dollar outlay, replace variables X_1 and X_3 which are most significant in the case of Y_1. Size of household (X_7) is important in two of the areas and length of occupancy (X_{10}) is important in three. However a new variable found important at this level is the age of the head of the household (X_{11}).

A negative relationship is noted between the age of the householder and the distance travelled.

4.5.3 Frequency of Trip (Y_3)

The range in the values of the multiple correlation coefficient is from 0.317 in Area 1 to 0.602 in Area 2 (see Table IV-4). In no case does the model explain more than one third of the variation in the number of visits made to grocery stores by the sample consumers. The equations derived from the stepwise program for the areas are set out below:

Area 1 $2.64 + 3.66X_3$

Area 2 $2.81 + 2.29X_1 + 4.80X_7 - 5.85X_9$

Area 3 $2.89 + 2.51X_1 + 1.41X_6$

Area 4 $2.56 + 2.08X_1$

Area 5 $3.03 + 1.44X_7 + 6.48X_1$.

As in the case of Y_1 and Y_2, the most important variables explaining variation between consumers on Y_3 are modes of travel. It is interesting to note that the size of family (X_7) is found to be important in Area 5, where the R^2 value is the second highest. Also note the negative relationship in Area 2 between income and frequency of trip.

4.6 Third Scale of Analysis

A third scale of spatial aggregation is isolated, and used for analysis. The areal unit in this case is the enumeration area. These

areal units contain about 150 households and cover approximately six city blocks. They are the smallest units used in census data collection. The sample area consists of 17 such enumeration areas. Again it is hypothesised that the relationships found at the first and second levels of analysis are not necessarily the same at this level, but vary according to the attributes of the smaller units. This analysis is instrumental in isolating the complexity of the relationships at this fine level of analysis.

4.6.1 The Means and Standard Deviations

The means and standard deviations of all the variables for the enumeration areas are given in Tables IV-5 to IV-8. The mean of the mode of travel variable X_1 varies from 0.04 in Area 158 to 0.69 in Area 173. This is understandable when the relative locations of the two areas are considered. Area 158 is located in the extreme south of the sample area and there is only one grocery store within one third of a mile from the area. Area 173 on the other hand is located on Concession Street where three grocery stores are located within the unit itself. Hence in the former most consumers travel by car to the grocery stores, while in the latter they walk.

This situation is also reflected in the distance variables, whereby the longest distances recorded are found in the enumeration areas where the car is commonly used to make the trip to the grocery store. Low mean distances are recorded for several areas, viz. Areas 162, 165 and 166 where both convenience grocery stores, and a supermarket are located either close to the areas, or within them. The

TABLE IV-5

Variable	ENUMERATION AREA 158		ENUMERATION AREA 159		ENUMERATION AREA 160		ENUMERATION AREA 161	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
X_1	0.04	0.19	0.05	0.24	0.27	0.45	0.22	0.43
X_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
X3	0.96	0.20	0.94	0.24	0.73	0.45	0.77	0.43
X_4	0.04	0.20	0.00	0.00	0.04	0.20	0.22	0.43
X_5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
X_6	0.96	0.20	1.00	0.00	0.96	0.20	0.77	0.43
X_7	2.64	0.81	3.00	0.68	2.62	0.64	2.56	0.70
X_8	3.36	0.99	3.44	0.98	3.46	1.03	3.17	1.15
X_9	3.24	1.13	3.44	1.15	3.33	0.89	3.00	0.91
X_{10}	4.20	1.12	3.89	1.28	4.00	1.06	4.00	0.84
X_{11}	3.48	0.92	3.44	1.10	3.31	0.97	3.56	1.04
Y_1	3.08	1.29	3.28	1.81	2.31	1.44	2.06	1.30
Y_2	3.28	1.27	3.78	1.89	2.73	1.40	2.22	1.26
Y_3	2.32	1.52	2.61	1.82	2.96	1.56	2.06	0.87

NOTE: Reference should be made to the questionaire (see Appendix) for interpretation of the means and standard deviations listed above.

TABLE IV-6

VARIABLE	ENUMERATION AREA 162		ENUMERATION AREA 163		ENUMERATION AREA 164		ENUMERATION AREA 165	
	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.
X_1	0.36	0.49	0.50	0.51	0.23	0.44	0.50	0.52
X_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
X_3	0.64	0.49	0.50	0.51	0.77	0.44	0.50	0.51
X_4	0.18	0.39	0.20	0.41	0.08	0.28	0.28	0.47
X_5	0.00	0.00	0.05	0.32	0.00	0.00	0.14	0.36
X_6	0.82	0.39	0.75	0.44	0.92	0.28	0.57	0.51
X_7	2.68	0.65	2.95	0.78	2.92	0.49	2.57	0.65
X_8	3.59	0.67	3.38	1.19	3.31	1.11	3.36	1.08
X_9	3.05	0.84	3.02	0.97	3.46	0.97	2.79	1.19
X_{10}	3.91	1.30	3.77	1.51	4.38	1.19	4.21	1.19
X_{11}	3.45	0.91	3.38	1.31	4.08	0.86	4.21	0.89
Y_1	2.14	1.58	2.05	1.79	2.08	1.85	1.64	1.34
Y_2	2.59	1.71	3.00	1.95	3.23	2.13	2.07	1.54
Y_3	3.32	1.84	2.87	1.60	2.85	1.82	3.29	1.20

NOTE: Reference should be made to the questionnaire (see Appendix) for interpretation of the means and standard deviations listed above.

TABLE IV-7

VARIABLE	ENUMERATION AREA 166		ENUMERATION AREA 167		ENUMERATION AREA 168		ENUMERATION AREA 169	
	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std. Dev.
X_1	0.50	0.67	0.19	0.40	0.46	0.52	0.50	0.52
X_2	0.45	1.47	0.00	0.00	0.00	0.00	0.06	0.22
X_3	1.05	1.36	0.76	0.44	0.53	0.52	0.45	0.51
X_4	0.14	0.35	0.05	0.22	0.20	0.41	0.25	0.44
X_5	0.09	0.29	0.00	0.00	0.00	0.00	0.05	0.22
X_6	1.23	0.92	0.95	0.22	0.80	0.42	0.70	0.47
X_7	2.45	1.06	3.10	0.83	2.66	0.62	2.35	0.74
X_8	3.23	1.57	3.05	0.92	2.80	1.15	3.25	1.37
X_9	3.00	1.20	3.43	1.08	2.93	1.16	3.20	1.20
X_{10}	4.14	1.55	4.29	1.01	4.13	0.99	4.40	1.18
X_{11}	3.73	1.55	3.67	0.97	3.80	1.01	4.20	1.15
Y_1	1.50	1.10	2.90	1.81	2.07	1.33	2.45	1.82
Y_2	1.64	1.26	3.33	1.74	3.00	1.65	2.55	1.67
Y_3	2.41	1.68	2.86	1.79	2.67	1.63	2.40	1.35

NOTE: Reference should be made to the questionnaire (see Appendix) for interpretation of the means and standard deviations listed above.

TABLE IV-8

Variable	ENUMERATION AREA 170		ENUMERATION AREA 171		ENUMERATION AREA 172		ENUMERATION AREA 173		ENUMERATION AREA 174	
	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev,	Mean	Std Dev.
X_1	0.53	0.52	0.42	0.51	0.25	0.45	0.69	0.48	0.50	0.51
X_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
X_3	0.46	0.52	0.58	0.51	0.75	0.45	0.31	0.48	0.46	0.51
X_4	0.40	0.51	0.26	0.45	0.19	0.40	0.38	0.50	0.38	0.49
X_5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
X_6	0.60	0.51	0.74	0.45	0.81	0.40	0.63	0.50	0.58	0.50
X_7	2.73	0.88	2.32	0.75	2.75	0.86	2.81	0.98	2.63	1.13
X_8	3.27	0.80	3.26	1.37	3.37	1.36	3.31	1.35	2.50	1.56
X_9	3.20	1.32	3.11	1.52	2.94	1.12	3.38	1.54	3.00	1.77
X_{10}	4.13	1.25	4.37	1.30	3.75	1.57	4.00	1.21	3.58	1.50
X_{11}	3.87	1.30	4.47	1.47	3.44	1.03	4.00	1.26	3.96	1.60
Y_1	1.33	0.62	2.74	1.96	2.94	2.08	2.13	1.75	1.79	1.74
Y_2	2.33	1.72	3.21	1.93	2.38	1.99	2.32	1.79	2.25	2.05
Y_3	2.87	1.68	2.68	1.66	2.38	1.26	3.44	1.67	3.04	1.68

NOTE: Reference should be made to the questionaire (see Appendix) for interpretation of the means and standard deviations listed above.

range in the mean of the "frequency of trip" variable (X_{14}) is from 2.06 in Area 161 to 3.44 in Area 173. The higher values are probably due to a number of stores being located within walking distances of consumers' residences.

4.6.2 A Cautionary Note

Data for the distance variables are in classes and this presents a problem which should be noted, especially at this finer scale of analysis. It is found that in some of the estimating equations the pure constant is the only significant estimate of the independent variable. In these cases the most efficient estimate of Y is given by $\overline{Y}$. The variance around $\overline{Y}$ is minimized because the data have been put into classes. Equations where only the pure constant is included emphasises that the absolute location of the areal unit in relation to the outlets those consumers travel to accounts for the most variance. Hence the appearance of only a pure constant in the estimating equations stated below is indicative of strong external locational forces affecting consumer spatial behavior in the area. Strong external locational forces means the nearest grocery store is located at a substantial distance from the area being analysed. The distance to that store therefore has a strong influence on the minimum distance consumers must travel to make grocery purchases.

4.6.3 Distance Measure I (Y_1)

The multiple regression correlation coefficients on Y_1 for the 17 enumeration areas range from 0.5451 to 0.9934 (see Table IV-9). The

TABLE IV-9

STEPWISE MULTIPLE REGRESSION ON ENUMERATION AREAS

Enumeration Area	Multiple "r" for Dist. I	Multiple "r" for Dist. II	Multiple "r" for Frequency
158	0.4969 (25%)	0.4969 (25%)	0.6024 (36%)
159	0.7650 (59%)	0.7289 (53%)	0.5251 (28%)
160	0.6961 (48%)	0.6405 (41%)	0.3380 (12%)
161	0.6956 (48%)	0.5316 (28%)	0.5508 (30%)
162	0.7505 (56%)	0.7605 (57%)	0.6849 (47%)
163	0.6705 (45%)	0.7001 (49%)	0.7651 (58%)
164	0.7372 (54%)	0.8415 (71%)	0.7792 (61%)
165	0.6848 (47%)	0.7755 (60%)	0.8984 (81%)
166	0.5451 (30%)	0.7328 (54%)	0.8952 (81%)
167	0.6725 (45%)	0.4471 (20%)	0.7647 (58%)
168	0.7963 (63%)	0.6863 (47%)	0.6347 (40%)
169	0.7916 (63%)	0.6374 (41%)	0.7524 (57%)
170	0.8119 (66%)	0.9059 (82%)	0.8440 (71%)
171	0.7879 (62%)	0.7241 (52%)	0.7827 (61%)
172	0.7532 (57%)	0.8169 (67%)	0.7146 (55%)
173	0.9934 (99%)	0.9693 (94%)	0.6591 (43%)
174	0.6962 (48%)	0.6869 (47%)	0.7427 (55%)

Figures in parentheses represent the percentage of the dependent variable which is explained by the independent variables.

highest values of "r" are obtained in the northern enumeration areas. The average explanation for the 17 areas is 55% which is a good indication that the variables are of more use in predicting the values of the dependent variable at this finer level of analysis than when the entire sample area was considered. This is largely due to the fact that the internal locational differentials in a small unit are smaller than in a large unit. In Area 173, where the external locational influences are weak, almost all the variation in the dependent variable is estimated in terms of the chosen independent variables. The estimating equations obtained, including only those variables which are most significant, are as follows:

Enumeration Area	158	$Y_1 = 3.08 + 292X_4$
	159	$Y_1 = 3.27$
	160	$Y_1 = 2.31 + 2.74X_1 + 4.59X_{10}$
	161	$Y_1 = 2.05 + 1.26X_9$
	162	$Y_1 = 2.31 + 2.78X_1 - 2.52X_8$
	163	$Y_1 = 2.05 + 3.10X_1 + 4.86X_7$
	164	$Y_1 = 2.08$
	165	$Y_1 = 1.64 + 2.28X_1$
	166	$Y_1 = 1.50$
	167	$Y_1 = 2.90 + 1.40X_3$
	168	$Y_1 = 2.06 + 2.87X_1$
	169	$Y_1 = 2.39 + 1.27X_3$
	170	$Y_1 = 1.33 + 1.71X_1 + 2.82X_{10}$
	171	$Y_1 = 2.73 + 3.63X_1 + 7.08X_7$

Enumeration Area 172	$Y_1 = 2.93 + 3.50X_1$
173	$Y_1 = 2.13 + 4.60X_1 + 4.17X_8$
174	$Y_1 = 1.79 + 1.00X_3$

As in the case with the analysis at the other levels of aggregation, mode of travel proves to be the most important variable in explaining the variation in the distance to the store which is visited most frequently by consumers to make grocery purchases. Socio-economic characteristics are found to be important in seven enumeration areas. Length of occupancy (X_{10}) is found to be important in Areas 160 and 170, the number of persons in the household (X_7) in Areas 163 and 171, the education of the head of the household (X_8) in Areas 162 and 173, and the income of the household (X_9) is the most important variable in Area 161. Education and income do not appear as significant variables at the other two levels of analysis.

4.6.4 Distance Measure II (Y_2)

The multiple regression coefficients with respect to Y_2 range from a low of 0.4771 for Area 167, to a high of 0.9693 for Area 173 (see Table IV-9). The average explanation for the areas is 53%, a figure which compares favorably with that obtained for the first distance measure. It is noted that in six of the areas $\overline{Y}_2$ is the only significant estimate parameter. This is explained by the reasoning set out earlier (see 4.6.2). The estimating equations derived from the regressions are listed below.

Enumeration Area 158	$Y_2 = 3.28$
159	$Y_2 = 3.36$
160	$Y_2 = 2.73$
161	$Y_2 = 3.37$
162	$Y_2 = 2.59 + 1.13X_3$
163	$Y_2 = 3.65 + 3.00X_4 + 1.15X_7$
164	$Y_2 = 3.23 + 8.60X_{10}$
165	$Y_2 = 1.64$
166	$Y_2 = 1.60 + 2.05X_1$
167	$Y_2 = 3.33$
168	$Y_2 = 2.05 + 1.33X_6$
169	$Y_2 = 2.55 + 1.81X_3$
170	$Y_2 = 2.33 + 3.11X_4 + 2.57X_1 + 4.84X_{10}$
171	$Y_2 = 3.21 + 1.40X_6$
172	$Y_2 = 3.37 + 6.79X_9$
173	$Y_2 = 2.31 + 4.60X_1 + 6.03X_{11}$
174	$Y_2 = 4.90 + 2.83X_6$

Mode of travel variables are again found to be the most important except in Areas 164 and 172 where length of occupancy (X_{10}) and income of the household (X_9) are found to be the most important variables. The number of persons in the household (X_7) is found significant in Area 161, while in Area 173 the age of the head of the household (X_{11}) appears as a significant variable for the first time at any scale of analysis.

4.6.5 Frequency of Trip (Y_3)

The range in the values of the multiple regression coefficients is greatest for Y_3. The lowest value obtained is 0.3380 in Area 160 and the highest is 0.8984 found in Area 165 (see Table IV-9). The average explanation is 52% which, in comparison with the values obtained at the first and second levels of analysis, is a large improvement. It is noticeable that the R^2 value obtained for the areas in the extreme south of the sample area are low, while those in the central and northern areas are much higher. The most significant contributions made to the R^2 values are shown in the estimating equations below:

Enumeration Area 158	$Y_3 = 2.32 + 6.00X_3$
159	$Y_3 = 2.61$
160	$Y_3 = 2.96$
161	$Y_3 = 3.07$
162	$Y_3 = 3.31 + 5.97X_{10} + 5.58X_1$
163	$Y_3 = 2.87 + 2.15X_1 + 7.74X_7 + 1.73X_8$
164	$Y_3 = 2.84 - 3.00X_7$
165	$Y_3 = 3.28$
166	$Y_3 = 2.41 + 3.22X_3 + 1.54X_6 + 2.95X_9$
167	$Y_3 = 2.85$
168	$Y_3 = 2.66$
169	$Y_3 = 1.02$
170	$Y_3 = 2.86$
171	$Y_3 = 2.68 + 2.70X_7 + 2.30X_9$
172	$Y_3 = 2.37 + 2.00X_1$

Enumeration Area 173	$Y_3 = 5.36 + 2.01X_8$
174	$Y_3 = 3.04 + 2.41X_1$

Size of family (X_7), education (X_8), income (X_9) and length of occupancy (X_{10}) are all important variables in various enumeration areas and in some cases they are the most important variables. Note that in eight areas absolute external location is very important (i.e. $Y_3 = \overline{Y}_3$).

4.7 Summary

In this chapter the results from the various regression models are summarized. Analysis is carried out at three different levels of spatial aggregation, and the most important variables are isolated using the stepwise compilation of a multiple regression program. An attempt is made in the next chapter to explain the reasons for the differences from one scale to another.

CHAPTER V

INTERPRETATION AND IMPLICATIONS OF RESULTS

The purpose of this chapter is to analyse the differences in the results for the different scales of analysis. One of the ways in which this is carried out is via the socio-economic variations in the compositions of the areas. Also, an exploratory analysis of the residuals from the regression models is made for sample areas at two scales of analysis. The residuals can be interpreted as reflecting the effect of locational differentials in the population, plus a "noise" component consisting of the omitted psychological variables, other socio-economic variables and elements which may be considered as random. It is hypothesised that the residuals reflect the spatial arrangement of variables not included in the regression models. The approach seeks to explain some of the differences for each dependent variable at each of the different scales.

5.1 Distance Measure I (Y_1)

The differentiation within the entire sample area is measured by the coefficient of variation for the variables used in the model. This is a measure which expresses the magnitude of the variation relative to the size of whatever is being measured. It is a standardized measure which is useful in comparing the variation among different areas, and

for t! same area at different scales (see Table V-1).

TABLE V-1

COEFFICIENTS OF VARIATION IN THE ENTIRE SAMPLE AREA

Size of household	X_7	30%
Education of head of household	X_8	30%
Income of household	X_9	36%
Length of occupancy	X_{10}	30%
Age of head of household	X_{11}	31%
Distance Measure I	Y_1	73%
Distance Measure II	Y_2	63%
Frequency of Trip	Y_3	57%

All of the socio-economic independent variables show approximately the same amount of variation within the sample area. At the second scale of analysis, the greatest variation among the attributes is found in Areas 4 and 5 (see Table V-2). These are also the areas in which the R^2 values are the highest for Y_1.

The picture is more complex using enumeration areas at the third level of analysis. The contrasts in variation are in some cases much greater than at the other two scales and in some cases much less (see Table V-3). For example in Area 164 the average coefficient of variation for the five socio-economic variables is 25%, while in Area 174 it is 46%. The areas in which the model is most successful are found in the northern part of the sample area: those areas adjacent to Concession

Street which contain a variety of retail grocery outlets.

TABLE V-2

COEFFICIENTS OF VARIATION AT THE SECOND SCALE OF ANALYSIS

		AREA 1	AREA 2	AREA 3	AREA 4	AREA 5
Size of family	X_7	27	27	26	30	38
Education	X_8	29	31	33	38	42
Income	X_9	31	30	34	38	51
Length of occupancy	X_{10}	28	34	23	31	35
Age	X_{11}	29	33	24	30	35
Distance Measure I	Y_1	53	78	78	74	83
Distance Measure II	Y_2	48	64	68	63	76
Frequency of Trip	Y_3	61	57	55	57	55

From an examination of the estimating equations, it is found that generally, mode of travel is the most important factor. However in Area 161, income of the household (X_9), is the most important and the education of the head of the household (X_8) is the next important variable.

5.2 Distance Measure II (Y_2)

The mean distance a consumer travels to the grocery store where he spends the largest amount of money on a single trip is longer at all three scales of analysis than the distance a consumer travels to the grocery store he visits most frequently. The degree to which $\overline{Y}_2$ exceeds $\overline{Y}_1$ is dependent on the locational attributes of each area.

TABLE V-3

COEFFICIENTS OF VARIATION FOR THE ENUMERATION AREAS

Enumeration Area	Size of Household	Education	Income	Length of Occupancy	Age of Householder	Distance Measure I	Distance Measure II	Freq. of Trip
158	30	29	33	26	26	42	39	66
159	23	28	33	33	32	55	50	70
160	24	30	27	26	29	62	51	52
161	27	36	30	21	29	63	57	42
162	24	19	28	33	26	74	66	55
163	27	33	32	40	39	87	64	56
164	17	34	28	27	21	88	65	64
165	25	32	43	28	21	81	74	36
166	43	49	40	38	42	73	76	69
167	27	30	31	23	26	62	52	63
168	30	36	32	25	30	64	54	61
169	32	42	37	27	27	74	65	56
170	32	24	41	30	34	46	73	58
171	32	42	49	30	33	72	60	62
172	31	40	38	41	30	70	59	52
173	35	41	46	30	32	82	76	46
174	43	45	59	42	40	99	91	55

For the entire sample $\bar{Y}_2$ exceeds $\bar{Y}_1$ by only one fifth of a mile. This is due mainly to the compensatory effects of both near and distant consumers frequenting a wide range of grocery outlets. This is similar to the situation shown in Figure 1.

For the final area at the second level of analysis the $\bar{Y}_2$'s exceed the $\bar{Y}_1$'s by distances ranging from one seventh of a mile to one quarter of a mile.

At the third scale of analysis the $\bar{Y}_2$'s exceed the $\bar{Y}_1$'s by distances ranging from as little as one twentieth of a mile in Area 169 to almost one half mile in Area 164.

Extremely long distances must be travelled in some areas, e.g. Area 172, and short distances in others, e.g. Area 166. Thus the conclusion that the largest purchase trip is longer than the most frequent trip during a week is a relatively scale free conclusion.

Also in most cases the regression model is more successful in explaining Y_1 then Y_2. This is especially true at the first and second levels of analysis. At the third level of analysis, ten out of the seventeen areas have higher R^2 values for Y_1 than for Y_2. A block of areas (162, 163, 164, 165 and 166) shows a higher R^2 for the Y_2 variable. The only conclusion which can be made here is that the model is more successful in predicting short distances, than the longer distances encountered for Y_2.

A correlation between the ranked R^2 values for Y_2 and the average coefficient of variation for the five socio-economic variables (X_7 to X_{11}) gave an "r" of 0.71. Thus in the areas where the model is more

successful greater variances in the attributes of the population are found. Mode of travel is generally the most important variable at all the scales, although at the third scale of analysis some of the socio-economic variables have local dominance. For example in Area 172, income (X_9) is the most important factor contributing to variation in the dependent variable. The number of persons in the household (X_7) is an important factor in Areas 163 and 171 and length of occupancy is important in Areas 160 and 170.

5.3 Frequency of Travel (Y_3)

A negative relationship exists between frequency of trip and the distance travelled to the most frequently visited store (Y_1). The variables found to be the most important are mode of travel, (X_1 - journey made on foot to the most frequently visited store) and the number of persons in the household as hypothesised. The first variable reflects the number of times that a family runs out of those common grocery items which are required at frequent intervals. With a greater number of people in the household, more individuals may be available to make trips.

At the second level of analysis the same picture is revealed, but local differences should be noted. In Area 1, journey made by car to the most frequently visited store (X_3), is more significant than walking to that store because consumers in this area have to travel longer distances to make grocery purchases. In Area 5, where distances are generally shorter, the mode of travel variable is replaced by the size of the family (X_7) in importance.

At the third level of analysis, the situation is more complex. The estimating equations are not as useful because of the reduced number of degrees of freedom. Of the nine enumeration areas where significant variables are included in the estimating equation, five different variables are found to be of prime importance locally. Besides mode of travel variables, and the size of the household variable (which are important at the other two levels), education and length of occupancy are found in the estimating equations.

5.4 Investigation of Residuals

In Chapter I it is noted that consumer spatial behavior is the outcome of a complex mesh of socio-economic, locational and psychological differentials in the population, and that it is most probable that the strength with which these factors affect consumer spatial behavior varies according to the scale at which analysis is made. From the preceding analysis it has been shown that the socio-economic and mode of travel set of variables are of increasing use as predictor elements, as the sample unit chosen as the base for analysis decreases in size.

Part of this increase may be attributable to the smaller sample sizes used at the enumeration level of analysis, but part of it may be due to the fact that at the finer level of analysis, internal locational differentials are much less important. Hence the socio-economic variables are of more use in explaining differences among the travel behavior of individuals.

This hypothesis is tested by plotting the residuals from regressions carried out at different scales. The first and the third scale

are taken since they approximate the hypothetical situations outlined in Figures 1 and 2 in Chapter I. Locational patterns may show up in the maps of residuals at one scale and not at the other.

The area selected for an analysis of residuals incorporates enumeration Areas 161, 162 and 163. The residuals from the regression on the entire sample, and the residuals from the separate enumeration area regression models are plotted for the two distance variables.

5.4.1 Distance Measure I (Y_1)

The residuals plotted from the individual enumeration area regressions are much smaller than those from the regression on the entire sample. This is a direct result of the difference in the R^2 value calculated for the regressions. It is difficult to recognize a pattern in the spatial distribution of the residuals. There is no area where there are groups of large positive or negative residuals. Large residuals tend to occur only as isolated cases, representing households who do not conform to the general pattern for some as yet unexplained reason. If these large residuals are the outcome of locational differentials among the population, then one would expect them to be clustered. The most probable reason why no pattern appears in the residuals for Y_1 is that there is a fairly even distribution of grocery stores over the entire area and so there is little opportunity for locational differentials to exert themselves.

5.4.2 Distance Measure II (Y_2)

The residuals from the first and third scales of analysis of

the sample area are shown in Figures 9 and 10. The distance in the vast majority of cases represents distance to a supermarket (stated earlier in Chapter IV). Since this type of outlet occurs less frequently on the landscape, locational differentials are more likely to be visible from a mapping of residuals. At the level of the enumeration area, the residuals do not show a marked pattern while in an analysis of the entire sample area the residuals do show a marked locational orientation. The major store in the chosen sample area is the A & P Supermarket located in enumeration Area 163 (see Figures 9 and 10). The highest residuals are found close to this supermarket. This would suggest that at the macro scale of analysis, locational differentials among the population are evident from the regression using only socio-economic and mode of travel variables. These locational differentials are not as evident when using the enumeration area as the base for the regression model. Hence it is justifiable to conclude that the locational differentials are more important in an analysis of the entire sample area, than when a small areal subset is the base for analysis.

Scale has been important in isolating the importance of the different sets of variables. Both socio-economic and locational differentials explain different amounts of variation in the dependent variables at different scales. Also different socio-economic variables are important at different scales.

5.5 Limitations of the Empirical Test

Five limitations are noted. The empirical test implemented in this thesis is purely exploratory in nature. Only three scales of analy-

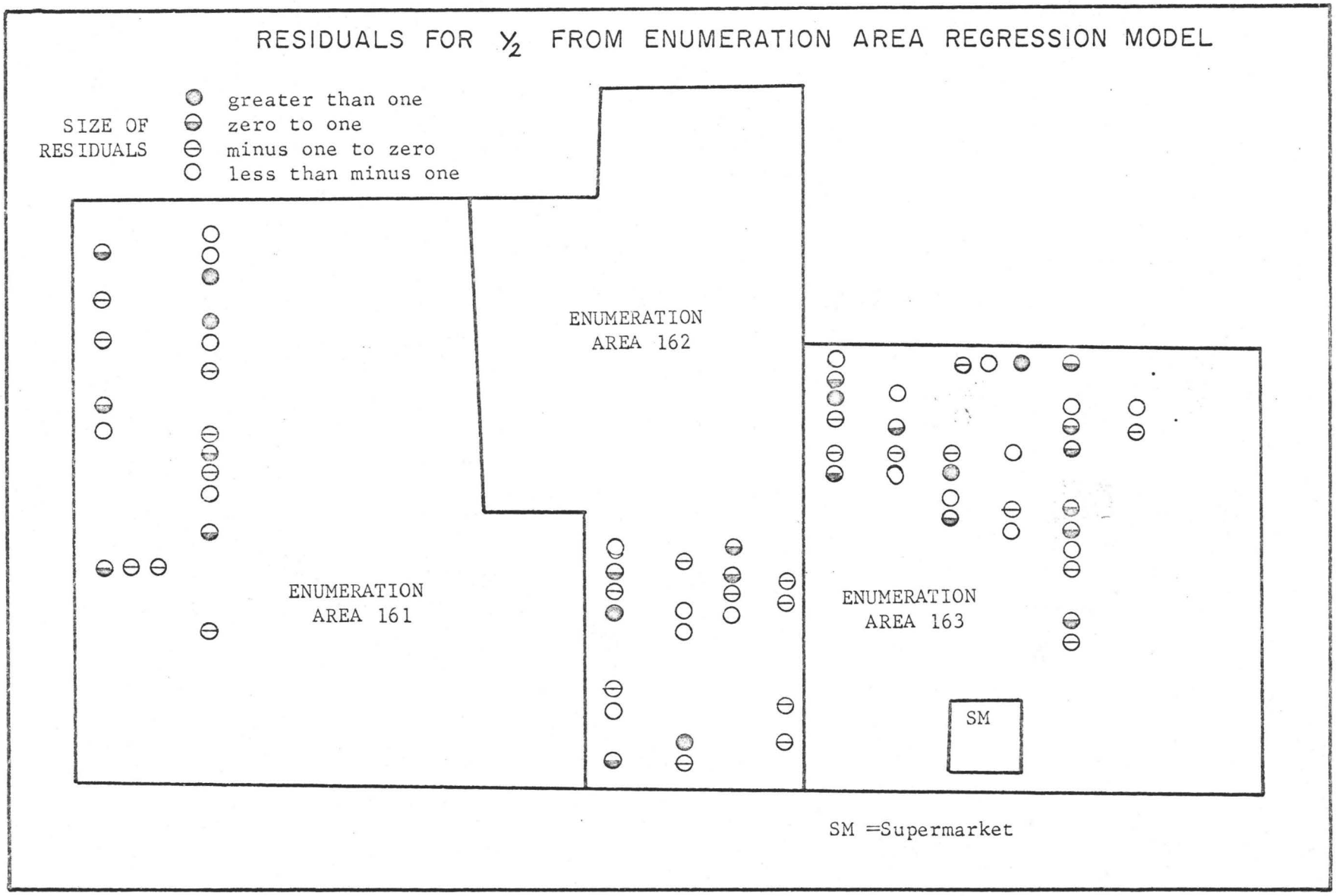

RESIDUALS FOR Y/2 FROM ENUMERATION AREA REGRESSION MODEL
SIZE OF RESIDUALS
greater than one
zero to one
minus one to zero
less than minus one
ENUMERATION AREA 161
ENUMERATION AREA 162
ENUMERATION AREA 163
SM
SM =Supermarket

FIGURE IX

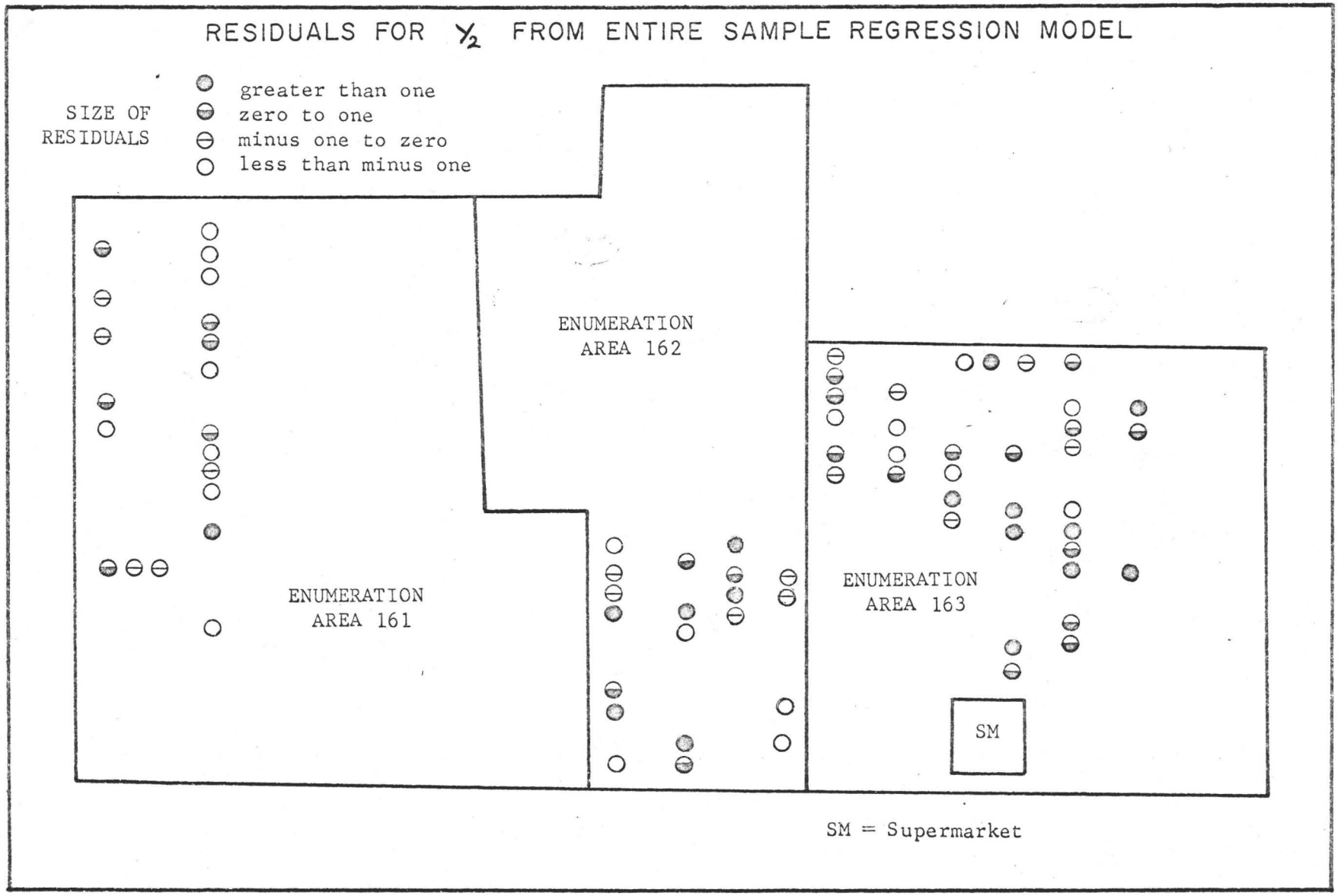

RESIDUALS FOR Y_2 FROM ENTIRE SAMPLE REGRESSION MODEL
SIZE OF RESIDUALS
greater than one
zero to one
minus one to zero
less than minus one
ENUMERATION AREA 161
ENUMERATION AREA 162
ENUMERATION AREA 163
SM
SM = Supermarket

FIGURE X

sis are taken as a basis on which the regression models are applied. Also the three scales selected are near the extreme lower end of a continuum of possible scales for analysing intra-urban phenomena. However, they are scales relevant to an entrepreneur interested in finding out important predictive relationships concerning consumer spatial behavior.

The sample sizes are small because of the restraints on time and resources available. Since random clustered samples are taken instead of the more commonly used random sample of individuals, each sample may not be as truly representative of the area as a random sample. However for the purposes of testing the main hypothesis of the study, the random clustered sample is considered adequate.

In the questionnaire analysis, data are only collected on the socio-economic attributes of the sample population. However, in the latter part of the thesis certain inferences are made regarding the locational differentials among the population, and how these contribute to an explanation of consumer spatial behavior at different scales of analysis. It is realised that the approach used at this point, namely the mapping of residuals from a regression of contributary socio-economic factors, is by no means the best method of doing this. However the patterns which did emerge are consistent with the observations which are expressed earlier. A better method is to include pertinent locational attributes of each household in the regression model, or in a separate regression model.

As realized at the outset of the study, there are many factors which explain consumer spatial behavior for many of which no

measures are as yet available. Hence there is a large "noise" component in the final explanation. This noise component includes many random variables, and since little or nothing is known about them at this stage in research, they can only be considered scale free.

The problem of attaining normality in the data is encountered. However, the information in the variables necessitated the collection of much data in classes, e.g. income and education. Thus, recognizing the limitations of the nature of the data, inferences are made with caution throughout the study.

CHAPTER VI

CONCLUSIONS

Using regression models it has been possible to give a readily understood value to the complex interrelationships which exist among the variables explaining intra-urban consumer spatial behavior. The complexity of these relationships is shown for three separate scales of analysis. Although primarily only socio-economic variables are investigated locational differentials are incorporated in the analysis.

The stepwise compilation of the regression model is used to isolate the variables contributing the most to variation in the three dependent variables considered. Mode of travel is the most important variable in explaining both the distance travelled to various grocery outlets and the frequency with which trips are made to these outlets. However, the length of occupancy by the householder at his present address and the number of persons in the household are the most important variables in several local areas.

A general change in the variables which are significant in explaining variation in the dependent variables from one scale of analysis to another is not found. However, the local complexity in relationships is noted. This supports the conclusion that the relationships which are found for one area may not be true in another area.

The general hypothesis that the relative importance of the factors

affecting consumer spatial behavior varies according to the level of spatial aggregation at which analysis is carried out is substantiated. It is found that socio-economic variables are more useful in small spatial units (e.g. enumeration areas) than for larger units (e.g. two census tracts). It is proved that the changes in explanation are due partly to changes in the importance of locational differentials among the population at different scales. At the micro-scale internal locational differentials are small, that is,most people live in approximately the same location and have approximately the same purchase alternatives offered to them. At the macro-scale of analysis, internal locational differentials are large, hence the range and accessibility of outlets is much greater. The regression model based on socio-economic variables alone is less successful in the larger spatial units than for smaller units.

Different variables are significant in different areas, even at the same scale. These differences are attributable to variation in both the means and the coefficients of variation of either some or all of the variables considered at the different scales. Some relationships however are found to be scale free conclusions. For example, at all three of the scales investigated it is found that consumers travel a longer distance to the grocery store where they spend the most money on a single trip per week, than the distance they travel to the store they visit the most frequently.

An exploratory look at the residuals from the estimating regression equations reveals a locational pattern which is compatible with the ob-

servation that locational differentials are more important at the macro-scale of analysis than at the micro-scale. A more comprehensive analysis of residuals may serve to isolate other variables which might be important in accounting for the remaining variation in the dependent variable. Thus, it is empirically proven that relationships derived at one scale of analysis are not universally applicable. Differences in the strength of relationships do exist for different scales.

The limitations of the study have already been noted, however mention here is made of several extensions it would be interesting to investigate in further research. Instead of three scales of analysis, a continuum of scales would give a more comprehensive image of the change in the importance of the attributes of consumers. Also it would be interesting to test what affect scale has on the statement of relationships for commodities, other than grocery products. It is hypothesised that scale will be more important in these cases, since outlets dispensing higher order goods generally occur less frequently on the landscape. Also of interest, especially in choosing a scale to analyse the phenomena, are the periodicities in consumer spending for goods. However, more sophisticated methods of analysis than the linear regression model will be required to operationalize these extensions.

APPENDIX

McMASTER UNIVERSITY

HAMILTON, ONTARIO, CANADA

RESEARCH UNIT FOR URBAN STUDIES

May, 1969.

Dear Sir or Madam,

The following questionnaire has been designed and distributed as part of a research programme being carried out by a graduate student at McMaster University, in co-operation with the Research Unit for Urban Studies. The success of this project depends on your co-operation, which will be gratefully appreciated.

The questionnaire is brief, and it should not take more than two or three minutes to complete. (I would ask that you complete it after reading this letter, as I know myself how easy it is to forget about something as soon as it is left.) THE INFORMATION CONTAINED IN THE QUESTIONNAIRE WILL BE COMPLETELY CONFIDENTIAL, and will only be used after the data has been aggregated.

The questionnaire was deposited in your mail box by a member of the research team on . He shall return to pick up the completed form on between and . If you have any problems completing the form, he will endeavour to help you at this time.

Thanking you in advance for your kind co-operation in this study, I remain,

Yours very truly,

SF/rt
Enclosure

Samuel Fulton,
Research Unit for Urban Studies.

McMASTER UNIVERSITY

HAMILTON, ONTARIO, CANADA

RESEARCH UNIT FOR URBAN STUDIES

May, 1969

Consumer Movement Survey

Please place a check mark (✓) in the appropriate box where applicable.

(1) Give the NAME and ADDRESS (approximate) of the grocery store which you have visited most FREQUENTLY DURING THE LAST TWO WEEKS.

NAME ----------

ADDRESS ----------

(2) Are the majority of the journeys to the above store made

on foot

by bus

by car

(3) Give the NAME and ADDRESS (approximate) of the grocery store where you expended the LARGEST AMOUNT OF MONEY ON A SINGLE TRIP DURING THE LAST 7 DAYS.

NAME ----------

ADDRESS ----------

(4) Was the visit to the above store (in 3) made

on foot

by bus

by car

(5) How many times did you visti all grocery stores in THE LAST 7 days?

a) Once

b) Twice

c) 3 times

d) 4 times

e) 5 times

f) Greater than 5 times

(6) How many persons are there in this household?

a) 1 person
b) 2-3 persons
c) 4-5 persons
d) 6-9 persons
e) 10 or more persons

(7) What is the level of education attained by the head of the household?

a) None
b) Elementary
c) High School 1-3 years
d) High School 4 years (Grade 12)
e) High School 5 years (Grade 13)
f) University

(8) What is the approximate gross income of your family per annum? (Sum of wages and salary of the members of the household)

a) Below $3000
b) $3000 - $6000
c) $6000 - $9000
d) $9000 - $12000
e) $12000 - $15000
f) Over $15000

(9) How long have you lived at your present address?

a) Less than 1 year
b) 1-2 years
c) 3-5 years
d) 6-10 years
e) Greater than 10 years

(10) Age of the head of the household?

a) Under 25 years
b) 25-35 years
c) 35-45 years
d) 45-55 years
e) 55-65 years
f) Over 65 years

LIST OF REFERENCES

BERRY, B. J. L. (1967) Geography of Market Centres and Retail Distribution. Englewood Cliffs.

CARROTHERS, G. A. P. (1956) "A historical review of gravity and potential models", Journal of the American Institute of Planners, 22, pp. 94-102.

CHRISTALLER, W. (1966) Central Places in Southern Germany, Englewood Cliffs.

CLARK, W. A. V. (1968) "Consumer travel patterns and the concept of range", Annals of the Association of American Geographers, 58, pp. 386-396.

COCHRAN, W. G. (1953) Sampling Techniques, New York.

COLENUTT, R. J. (1969) "Linear diffusion in an urban setting", Geographical Analysis, 7, pp. 71-78.

CURRY, L. (1966) "A note on spatial association", The Professional Geographer, 18, pp. 97-99.

---------- (1967) "Central places in the random spatial economy", Journal of Regional Science, 7, pp. 217-238.

DAVIES, R. L. (1969) "Effects of consumer income differences on shopping movement behavior", Tijdschrift voor Econ. en Soc. Geografie, 60, pp. 111-121.

DUNCAN, O. D. et. al. (1961) Statistical Geography; Problems in Analysing Areal Data, New York.

FREUND, J. E. (1967) Modern Elementary Statistics, Englewood Cliffs.

GARRISON, W. L. et. al. (1959) Studies of Highway Development and Geographic Change, Seattle.

GOLLEDGE, R. G. and BROWN, L. A. (1967) Search, learning and the decision-making process", Geografiska Annaler, 49, pp. 116-123.

HAGERSTRAND, T. (1965) "A Monte-Carlo approach to diffusion", European Journal of Sociology, 6, pp. 43-67.

HAGGETT, P. (1965) Locational Analysis in Human Geography, London.

HAGOOD, M. J. and PRICE, O. D. (1952) Statistics for Sociologists, New York.

HARVEY, D. W. (1968) "Pattern, process and the scale problem in geographical research", Proceeding, Institute of British Geographers, 40, pp. 71-78.

HORTON, F. E. (1968) "Locational factors as determinents of consumer attraction to retail firms", Annals of the Association of American Geographers, 58, pp. 787-801.

HUFF, D. L. (1960) "A topographical model of consumer space preferences", Papers & Proc. of the Reg. Sci. Ass., 6, pp. 159-173.

---------- (1963) "A probability analysis of shopping centre trade areas", Land Economics, 53, pp. 81-90.

ISARD, W. (1956) Location and Space Economy, New York.

KENDALL, M. G. (1939) "The geographical distribution of crop productivity in England", Journal of the Royal Statistical Society, 102, pp. 21-62.

KING, L. J. (1969) Statistical Analysis in Geography, Englewood Cliffs.

KNOS, D. S. (1956) "The distribution of land values in Topeka, Kansas", University of Kansas, Bureau of Business and Economic Research, pp. 15-33.

MARBLE, D. F. (1959) "Transport inputs at residential sites", Papers & Proc. of the Reg. Sci. Ass., 5, pp. 253-265.

McCARTY, H. H., HOOK, J. C. and KNOS, D. S. (1956) The Measurement of Association in Industrial Geography, Iowa City, Department of Geography, University of Iowa.

MURDIE, R. A. (1967) "Cultural differences in consumer behavior in Eastern Ontario", Canadian Geographer, 11, pp. 143-156.

NEFT, D. S. (1966) Statistical Analysis for Areal Distributions, Regional Science Research Institute, Monograph Series, Number 2.

NORBECK, S. (1963) "Computing distances in road nets", Papers & Proc. of the Reg. Sci. Ass., 12, pp. 207-230.

OLSSON, G. (1968) "Complementary models; a study of colonization maps", Geografiska Annaler, 50B, pp. 1-19.

RALSTON, A. and WILF, H. S. (1960) Mathematical Models for Digital Computers, New York.

ROBINSON, W. S. (1960) "Ecological correlations and the behavior of individuals", American Sociological Review, 15, pp. 351-357.

RUSHTON, G. (1966) "Spatial pattern of grocery purchases by the Iowa rural population", Studies in Business and Economics, New Series No. 9, University of Iowa.

RUSHTON, G., GOLLEDGE, R. G. and CLARK, W. A. V. (1967) "Formulation and test of a normative model for the spatial allocation of grocery expenditures by a dispersed population", Annals of the Association of American Geographers, 57, pp. 359-400.

SIEGAL, S. (1956) Nonparametric Statistics for the Behavioral Sciences, New York.

SUITS, D. B. (1957) "Use of dummy variables in regression equations", Journal of the American Statistical Association, 52, pp. 548-551.

THOMAS, E. N. (1968) "Maps of residuals from regression", in B. J. L. Berry and D. F. Marble (Eds.) Spatial Analysis, Englewood Cliffs, pp. 326-352.

THOMAS, E. N. and ANDERSON, D. L. (1965) "Additional comments on weighting values in correlation analysis of areal data", Annals of the Association of American Geographers, 55, pp. 492-505.

THOMPSON, D. L. (1963) "New concept - subjective distance", Journal of Retailing, 39, pp. 1-6.

TOBLER, W. (1966) "Spectral analysis of spatial series", Proceedings, Forth Annual Conference on Urban Planning Information Systems, Berkeley, pp. 179-185.

WARNTZ, W. (1965) Macrogeography and Income Fronts, Regional Science Research Institute, Monograph Series, Number 3.

YULE, G. U. and KENDALL, M. G. (1950) An Introduction to the Theory of Statistics, New York.