CONSUMER SPACE PREFERENCES

AN ANALYSIS OF CONSUMER SPACE PREFERENCES USING THE METHOD OF PAIRED COMPARISONS

By

GORDON ORR EWING, M.A. (Glasgow), M.A. (McMaster)

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

For the Degree

Doctor of Philosophy

McMaster University

September, 1971

DOCTOR OF PHILOSOPHY (1971) McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: An Analysis of Consumer Space Preferences using the Method of Paired Comparisons

AUTHOR: Gordon Orr Ewing, M.A. (Glasgow University)
M.A. (McMaster University)

SUPERVISOR: Dr. G. Rushton

NUMBER OF PAGES: X, 237

SCOPE AND CONTENTS:

The study seeks to derive a spatial preference model for urban places, based on a farm population's spatial choices of urban places for retail expenditure. Specifically, the study tests for the similarity in households' preference orderings of urban places, and finds a high degree of similarity. This is achieved using a model with only two simple variables, namely town population and distance to town. Tests indicate no major variable is omitted in the model. The information on households' preference orderings enables the aggregate preference order to be defined.

Tests are inconclusive as to whether households also assign worths to urban places which indicate their awareness of the amount by which different places are preferred.

The second part of the analysis seeks to determine if different types of households have different spatial preferences. Differences are revealed in the value different types attach to the locational convenience of places. Farm households with members working off the farm reveal a lower than average preference for convenient shopping

places. Smaller, more affluent households display a stronger preference for convenient locations and a lower preference for large towns than less affluent households with young children.

ACKNOWLEDGEMENTS

The writer wishes to thank Drs. Anderson, Carment and Rushton for their useful criticisms and encouragement. In particular I am grateful to Dr. Rushton of the University of Iowa who provided much of the stimulus for my interest in preference analysis. Also thanks are due to Mr. P. J. O'Day, Operations Supervisor of McMaster University's CDC 6400 installation, for so readily providing large amounts of computer time.

TABLE OF CONTENTS

Table of Contents (cont'd.)

List of Illustrations

List of Illustrations (cont'd.)

List of Tables

List of Tables (cont'd.)

INTRODUCTION

The research described here is based on certain premises, both philosophical and methodological, regarding human behavior and its measurement. The problem tackled is essentially one of attempting to define the form of spatial choice rules used by individuals to select amongst spatial alternatives. But before describing the problem in more detail, some explanation of the above-mentioned premises is in order.

Order in Human Behavior

Several approaches to the interpretability of human behavior are currently accepted. All approaches observe that individual behavior varies, i.e., that when faced with the same set of alternatives several times, people do not always make the same choice. However the two major approaches differ in their explanation of these variations. One assumes human behavior is a function of unpredictable random processes. The argument of this school of thought is that variation in behavior is not systematic and that explanations based on models of learning, adaptation, etc., have not and cannot adequately explain such variation. Presumably, students of the 'random process' school would deny the efficacy of the study of individual behavior and advocate instead a macro-analysis of human events which would relate aggregate patterns of human action to characteristics of the social, economic and physical environment.

The difference between this approach and the other is well summarised by Edwards and Tversky (1967, p. 315):

> Students of behavior disagree about whether this seeming randomness reflects actual random processes at work inside the organism, or whether it is the inevitable result of trying to use very simple methods of prediction to predict behavior that has very complicated causes. This disagreement is profound.

This other group of behavioral students who see the apparent randomness in behavior to be a consequence of our weak understanding of complex casual relationships adopt one of two approaches. One takes the approach that such is the complexity in relationships between factors affecting behavior that there is little value in attempting to establish rules of behavior at the level of the individual. Curry (1964, p. 138) epitomises this attitude:

> It is all very well when one can supply the parameters within which choices are made, but in any general location problem.....one cannot begin to comprehend the infinite number of decisions, rarely coincident in time and separately motivated under differing circumstances and degrees of information.

The other approach accepts the likelihood that complex relationships are the antecedents of behavior, but nevertheless seeks to discern order in behavior at the level of the individual since the individual is the basic unit of concern to the social scientist, and therefore insights gained at that level are the most useful. This is the approach adopted in the research described here.

Human Behavior, Choice and Preference

When an individual acts voluntarily he usually has accepted one course of action in preference to one or more alternatives. His action, therefore, can be regarded as the outward evidence of mental choice amongst alternatives. The basic premise of this study is that choice is not random but rather is based on some criterion used by the individual to assess the worth of different alternatives. The criterion can be thought of as the individual's assessment of which features of the alternatives are relevant to his choice, and of the importance of each. In this study such a criterion is referred to both as a choice rule and a preference function. Thus the choice of one alternative from among several would indicate that according to the criterion, the alternative chosen had more of the desired attributes than any other. In fact this is an oversimplification, for whilst the alternative may have most of one attribute, it may not have most of all relevant attributes. Thus in choosing between houses, one may be the best constructed but not be in the best neighborhood, so that it is not a simple matter of choosing the alternative which is best on all counts. From experience in choosing, we know that this is not an uncommon predicament. Presumably the choice of one alternative from amongst several depends both upon the quantity of each relevant attribute perceived in each alternative (q_{ij} = quantity of attribute i perceived in alternative j), and upon the individual's preferences rule, i.e., his weighting of each attribute by importance (w_i = weight attached to attribute i). Thus for a given alternative, each attribute would have a worth expressed as a function of its quantity and weighting denoted by v_{ij}, the worth of attribute i in alternative j. In addition, however, the individual presumably has some method of combining the worths of each attribute in an alternative (v_{ij}) by means of which he can assess the total worth of an alternative. Thus the overall worth of the alternative to an individual partly depends on how he combines the part-worths of each attribute. Thus,

$$V_j = f(v_{1j}, v_{2j}, \ldots\ldots, v_{nj}) \quad \text{(I.1)}$$

where

$$v_{ij} = f(w_i, q_{ij}).^{1} \quad \text{(I.2)}$$

The Rationale of a Behavior-Oriented Approach

The primary methodological premise implicit in the previous section is that a person's rule for choosing between alternatives - in a word, his preference - can be discerned from overt behavior involving choice amongst alternatives. Thus preference, a mental trait, is inferred from the overt behavior of a person choosing one alternative in preference to others. The rationale behind this behavioristic approach is well described by Bertrand Russell (1921, p. 47) who believes that:

> The discovery of our own motives can only be made by the same process by which we discover other people's, namely the process of observing our actions and inferring the desire which could prompt them.

Implicitly rejected as a method of determining preference is the mentalist approach whereby individuals are asked to verbally describe their preferences and "explain" or rationalise them. Miller (1962, p. 8) provides some explanation for this trend of behaviorism replacing mentalism:

---

[1] This is commonly known as the composition rule in psychology, and is discussed by, among others, Shepard (1964) and Tversky (1967) in its commonest form, the additive composition model.

> Standing opposed to this trend is the stubborn fact of consciousness; everyone feels that he has direct, immediate evidence concerning his own mind. However, a growing body of psychiatric and psychoanalytic experience argued that consciousness is too narrow a window to provide an unobstructed view of all that should be classed as mental. Consciousness may register the outcomes of thought, but the processes themselves remain hidden from our inner vision. Psychologists who tried to use scientific criteria and methods were forced more and more into the admission that they were studying behavior not consciousness.

Behavior, therefore, is the basis used to infer preference.

The Purpose of the Study

One purpose of this study, in the context of subjects choosing amongst multi-attributed spatial alternatives, is to determine the relative importance people attach to the attributes. Additionally it seeks to determine whether the relative importance of given quantities of each attribute is the same for different types of households, or if preference functions vary between household types, i.e., it seeks to determine the extent of inter-personal consistency in preferences.[2]

A third purpose, assuming orderliness is proven in the particular spatial preference functions considered here, is to determine whether individuals simply have a mental ordering of alternatives from most to least preferred, or if in fact they can more precisely locate alternatives on their mental preference scale. Chapter 1 defines these objectives

---

[2] To put the study in proper perspective, it should be noted that several other critical aspects of choice as defined in (I.1) and (I.2) above are ignored. Thus the extent to which an individual may vary his estimate of attribute importance (intra-personal consistency) is disregarded, as is the matter of intra- and inter-personal consistency in perception of the same alternative.

more specifically and explains the need for such analyses.

Chapter 1

THE PROBLEM

The general purpose of this study is to describe the nature and properties of the rule or rules by which consuming units choose between alternative central places for grocery purchases. Two specific aspects are considered. Firstly, to what extent do households reveal similar preferences between the same alternatives, and, if there are significant differences in preference, how large are they and how are they related to any social, economic or other behavioral characteristics of households? Secondly, do individuals simply rank spatial alternatives in order of preference or does their assignment of worths to alternatives indicate an awareness of quantities by which different alternatives are preferred to one another?

Inter-Personal Difference in Space Preferences

In light of the apparent success of some existing aggregative models in predicting individual choice behavior (Huff, 1962; Rushton, 1966), it might appear reasonable to infer that there is a considerable degree of agreement in individuals' space preferences, and little cause to hypothesize differences of any magnitude. However, to draw such a conclusion would be to ignore the possibility that the correctly predicted choice situations are of such a simple kind that many *different* models might predict such choices equally well. The real test of the generality of any preference model is when the individual is confronted with a complex choice situation, and until now the evidence has been that these are the situations where such models are most deficient (Rushton, 1966; Ewing, 1968). Such a deficiency may be explained, at least partly, by these models incorporating only a single preference function to predict the spatial choices of people with differing preference functions. In this research the hypothesis is tested that systematic differences exist in the spatial preferences of a sample of households with respect to grocery purchasing.

The following discussion provides a rationale for testing this hypothesis, by indicating possible sources of space preference differences, and by noting the frequency of this hypothesis in the literature.

Sources of Inter-Personal Variations in Preference and Suggested Explanations

The fact that two people differently evaluate the same alternative courses of action is attributable to three factors - the weight they attach to each attribute of the alternatives, the quantity of each attribute in each alternative as perceived by them, and their method of combining the part-worths of each attribute (v_i). Thus a person's choice rule is a function of his attribute weighting, of his perception, and of his method of combining the part-worths of attributes in a multi-attribute alternative. The introduction of the notion of perception requires that the original equation of part-worth,

$$v_{ij} = f(w_i, q_{ij}),$$

now be rewritten as

$$v_{ij} = f(w_i, q_{ij}, \alpha_i) \qquad (1.1)$$

where α_i = the perceptual weighting attached to a quantity of attribute i. It is conceivable that two individuals may attach precisely the same weightings to given attributes and yet make different choices from the same alternatives as a result of variations in their perception of the relative quantities of each attribute in each alternative.

The most common explanation of differences in perception of the same thing is that they are a function of differences in what has been learned about it. Various explanations of learning differences exist. For example, the more contact a person has with a set of spatial alternatives, the keener should be his awareness of differences between them. Thus Golledge (1967, p. 247) suggests that:

> In the course of satisfying such desires individuals will test a number of possible combinations of markets ... From their experience of the results they will 'learn' which decision process ... gives them the greatest rewards ... and they will tend to retain 'satisfactory' responses and delete 'unsatisfactory' responses. This process is likely to be continued as the search for the 'most satisfactory' pattern of responses is carried out.

The inference is that decision-making is sequential with the outcome from each decision contributing to the learning process. Individuals at different stages in learning the attributes of different spatial alternatives presumably may have a different perception of them.

A further explanation of differences in environmental learning, and hence in perception, is that persons from differing backgrounds tend to frequent different paths in the spatial network, thus introducing biases into their spatial information. Wolpert (1965, p. 165) suggests that:

> Differences in sex, race, formal education, family income and status are likely to find their expression in shaping the area of movement and choice. Although the action space is unique for each individual, still there is likely to be a good deal of convergence into a limited number of broad classes.

It would seem reasonable to hypothesize, therefore, that interpersonal preference differences may be a function of differing perceptions.

The second source of preference variation is that individuals may attach different weightings to the same attribute of an alternative. Thus, for example, the accessibility of a shopping centre may weigh more heavily with the unmarried working man than with the housewife, whilst she may place more value on the variety of goods offered at a retail outlet. Shepard (1964, p. 257) describes the problem succinctly in more general terms:

> The relative weights to be assigned to the component attributes are not always determinate and may, in fact, depend on the adoption of one of several incompatible but equally tenable systems of subjective goals ... (or 'state of mind').

It is also conceivable that less stark differences in preferences may exist than in the above example where accessibility is invariably more important to one individual than to another. Thus, for example, two individuals may agree upon the importance of accessibility of large urban centres, but vary in their value judgement of the accessibility of small centres. In other words, the weight attached to one attribute may vary between individuals only for a limited range of quantities of one or more attributes.

The third possible source of inter-personal variation in preference is the method of combining the part-worths of the different attributes relevant to choice. Different assumptions have been made regarding an appropriate form for the rules of combination. In psychology and economics, Edwards and Tversky (1967, p. 255) indicate:

> One idea so completely dominates the literature on riskless choice that it has no competitors. It is the additive composition notion. It asserts that the utility of a multi-dimensional alternative, such as a commodity bundle or a job offer, equals the sum of the utility of its components.

By contrast, in geography and marketing, the utility of a spatial alternative is commonly defined in terms of a gravity model where,

$$U_{ij} = M_i \times d_{ij}^{-\alpha} \qquad (1.2)$$

where α = an empirically derived constant,

U_{ij} = the utility of alternative i to an individual j,

M_i = some measure of the "mass" or attraction of alternative i,

and d_{ij} = the distance between i and j.

In this case the part-worths of mass and distance are multiplicative rather than additive.

Several possible sources of inter-personal variation in the components of a preference function have been indicated. This is reason enough to test for preference differences. Also, the test is prompted by the frequent discussion of the subject both in psychology and geography, together with an absence of rigorous tests in the geographic literature.

Interpersonal Preference Variation as Discussed in the Literature

The stimulus from the literature for a test of inter-personal differences in space preferences is two-fold. The combination of vast amounts of psychological research into inter-personal differences in intellect and attitude, and a dearth of rigorous analysis of preference differences by geographers provide adequate reason for such an investigation. In psychology the entire field of mental testing is designed to distinguish differences in people's intellectual, emotional and attitudinal traits. Nevertheless, relatively little investigation has been aimed at discriminating between individuals on the basis of their preferences. Thus Edwards (1961, p. 488) concludes:

> It is surprising that so few studies explicitly examine utility ... functions, relating their shape in different people to personality variables.

One of the few examples of an attempt to define individuals in terms of similarity of preference for the same set of alternatives is by Tucker (1960). Given each individual's ranking of the same set of alternatives, he shows a method conceptually akin to non-metric factor analysis, which defines each person's stimulus preference vector in a multi-dimensional preference space. He does not, however, pursue the topic in the manner suggested by Edwards, by trying to relate differences in their utility functions to personality traits, but simply concludes, not surprisingly, that individual preference differences do exist.

Geographers, in contrast to psychologists, have treated the notion of differing preference functions regarding spatial phenomena at a much less sophisticated technical and conceptual level. This point will be elaborated in Chapter 2. Nevertheless, the notion was discussed in the geographical literature as early as 1937 by Christaller (1966, p. 22):

> The range [of a good] is determined by ... .. the income conditions and the social structure of the population, ... and numerous other elements.

And elaborating on this in a footnote, he remarked that:

> Englander includes all of this under the term, 'price-willingness of the buyer'. This short expression is very striking; it means approximately that a certain population or a certain stratum of the population with regard to its structure and composition, is willing to pay a certain higher price for particular goods it desires. (1966, p. 26)

Subsequently the notion of differing space preferences has received frequent mention in the literature, but has been the subject of very few empirical analyses.[1] Besides references to it in central place works, differing space preferences have been discussed by Hagerstrand (1966) with respect to innovation diffusion, by Moore (1969) and Wolpert (1965) on migration, and by Michelson (1966) and Peterson (1967a, 1967b) with reference to urban residential preferences. In

---

[1] Those studies which present empirical tests for a relationship between personal characteristics and space preferences are reviewed in chapter 2.

central place works, the notion of differing space preferences has been considered in the works of Getis (1961), Huff (1959), Marble (1959), Mayfield (1963), Murdie (1965), and Ray (1967). The general tenor of these works is that personal characteristics probably influence spatial behavior, but that research on the subject is insufficient. Thus Marble (1959, p. 22) says of location theory in general that:

> while devoting little attention to the behavior of the individual decision-making unit, it does recognize the importance of spatial location as well as certain social and psychological factors (space preference) in determining individual behavior in space.

In contrast to the level of methodological refinement achieved by psychologists in the analysis of human preference, studies in geography have been somewhat circumscribed by methodological shortcomings which are discussed in Chapter 2. Likewise, geographical analysis of differing space preferences have been characterized both by methodological and technical weaknesses. Consequently, as a prerequisite to the initial purpose of testing for differing space preferences, this research must first develop a more refined methodology and improved technique to test spatial choice behavior for evidence of differing preferences.

Ordinal and Ratio-Scale Preference

The second major objective of the study is to determine what the derived preference structures reveal about the properties of the household's mental yardstick for comparing spatial alternatives. Specifically the question is whether people simply have a mental ranking of alternatives in order of preference, i.e. an ordinal preference scale, or whether they discriminate acutely enough between alternatives that each alternative can be assigned a specific quantity of "preferredness" in a ratio scale sense.[2] Such a scale would not only imply that preference is probabilistic, but that the probability of choosing one alternative from a set could be predicted in terms of the ratio scale scores of the alternatives in the set.[3] This would constitute not only interesting information about how individuals assign measures of importance to spatial alternatives, but as Luce (1959) has shown, such a metric scale would permit prediction of choice from amongst combinations of alternatives for which no direct observations of choice are available.[4] However, if the result of a test for a ratio scale of preference was negative, the implication would not necessarily be that preference is deterministic, but simply that the probabilities do not have the metric relationships necessary for a ratio scale. It would mean that, at best, no higher ordered preference metric than an ordinal one could be assumed. But the fact that an ordinal function permits only deterministic prediction of choices would not imply that the choices used to induce that ordering were necessarily deterministic.

In summary, there are at least three reasons to test for a ratio

---

[2] In psychology, the research on ratio scales tends to be confirmatory (Coombs, 1964, p. 365).

[3] This topic is treated in more detail in Chapter 4.

[4] Luce's contributions are discussed in Chapter 4.

preference scale. Firstly, such a scale would indicate a high sensitivity in people's discrimination between spatial alternatives. Secondly, it would enable more realistic probabilistic prediction of choice.[5] And thirdly, as Luce (1959) has indicated, a ratio preference scale permits the prediction of choices from amongst previously unrecorded combinations of alternatives.

The Study and Central Place Theory

Assuming that central place systems are in part a response to the pattern of consumer demand and to preferred interaction patterns, a test for significant differences in the preferences of different types of consumers constitutes, in effect, a test of whether consumer type is a significant independent variable affecting the size and spacing of central places.

The second part of the analysis concerned with the test for both ordinal and cardinal preference functions also has considerable relevance to central place analysis. Little research has been undertaken to discover the effect which an alteration of Christaller's basic consumer behavioral postulate (that people go to the nearest place supplying a desired good) would have on the pattern of a

---

[5]The validity of assuming probabilistic rather than deterministic preferences is discussed frequently in the psychological literature. For example, see discussions by Coombs (1964, pp 106-118), Clarke (1960) and Luce and Suppes (1965, p. 331).

central place system deduced using that postulate.[6] Christaller posits a normative rule of central place choice by the consumer, such that for any point on a demand surface and for any given good there is a unique centre which all consumers at that point will invariably patronize when that good is the highest order good required. An ordinal preference rule would posit precisely the same type of normative behavior. But as indicated earlier, probabilistic behavior is considered more realistic. In this respect, a ratio preference rule permits calculation of the probability of each centre being patronized by consumers at each point on a demand surface. A test is needed of whether the replacement of a deterministic preference rule by a probabilistic one is likely to significantly alter the properties of a central place system derived from it.[7]

---

[6] An exception is to be found in Rushton (1970). However, even here the effect of a different behavioral postulate is not tested by deducing a central place system using only consumer and retailer behavioral postulates as Christaller did in his original formulation. Rather, Rushton's original formulation includes not only a consumer behavioral postulate but also the special case of a k 3 central place hierarchy. Thus the test is of the impact of an altered behavioral postulate upon a "special case" spatial system, in which the subsequently derived central place system may or may not have the same parameters as if any other initial configuration had been posited.

[7] It should be noted in passing that certain problems arise in attempting to derive a central place system from a probabilistic preference rule, which do not exist with Christaller's deterministic rule, at least. In his work the main geometric problem is to pack an many circular trade areas into an area as threshold requirements permit, with the hexagonal solution resulting in a regular triangular lattice of centres. With other rules, either deterministic or probabilistic, the problem of arranging points in space to best satisfy consumer preference and retailer threshold needs is more complex. The problem is no longer the simple packing one associated with a distance minimising goal and is compounded by the likelihood of different preference functions for different commodities.

Certainly, at a local level the survival of a spatial alternative is likely to be affected not only by people's order of preference but also by whether consumers are almost indifferent between it and another alternative or almost invariably prefer it to the other alternative. A preference ranking lacks such information about the degree of preferredness between things. Thus, whilst the existence of not only an ordinal but a ratio preference scale would have intrinsic interest regarding the nature of human behavior, it would also be meaningful in terms of differences in the kinds of central place systems derived using an ordinal or ratio preference postulate.

The Study and Spatial Theory

Whilst the specific subject matter of consumers choosing amongst alternative places to shop is confined in relevance to central place theory, a broader view of the analysis suggests the pertinence of the subject matter to any aspect of location theory concerned with how individuals go about choosing from amongst courses of action which have spatial attributes relevant to the choice. Thus, for example, the industrialist's choice amongst potential plant sites, the migrant's choice amongst towns in a country or amongst houses in a town, and even the innovation's "choice" amongst potential adopters at each sequence in its diffusion, all involve analogous spatial decision-making processes with a change in actors rather than actions.

The two major problems have been described and supporting argument provided. Before describing the research design devised for the analysis, it is essential to demonstrate that the problems described have not been adequately studied in previous research. In Chapter 2 this theme is elaborated in a review and critique of the geographical literature describing spatial preference models.

Chapter 2 REVIEW OF SPATIAL PREFERENCE LITERATURE

Relatively few studies have sought to define empirical regularities in spatial preferences. Instead, classical works in location theory [von Thünen (1966), Weber (1928), Lösch (1954), Christaller (1966), Isard (1956), Alonso (1964)] and most derivative studies until recently have been based on normative and optimal preference assumptions. Weaknesses in these models are commonly attributed to differences between the model's environment and reality's, or to the omission of a time perspective in models not directly concerned with diffusion. Rarely is the assumption about the human agent's spatial preferences questioned. More recently some students of location theory have sought to define empirical preference regularities in order to provide location theories with more realistic postulates.

The review which follows concentrates on analyses which have sought to define realistic spatial preference rules or to identify personal and environmental characteristics affecting preference rules. Primarily, models relating to Central Place Theory are discussed, although others are mentioned. Within this framework, two types of preference analysis are reviewed separately, those concerned with a sample's spatial preferences and those concerned to explain differences in spatial preferences in terms of differences in consumer characteristics. A methodological weakness common to almost all the behavioral analyses is identified.

The Problem of Interdependency of Behavior and Environment

A basic problem in deriving consumers' preference rules is to obtain a rule which is in no way dependent on the particular arrangement of urban places in the study area used to derive the rule. Otherwise there might be almost as many rules as study areas. Curry (1967, p. 218) suggests this to be the major problem in writing central place theory, and therefore warns (1967, p. 219) that "a postulate of spatial behavior should not directly describe the behavior occurring within a central place system". In other words, it is necessary to avoid such things as inferring preference amongst spatial alternatives from the frequency with which different alternatives are chosen by consumers, unless the same alternatives are available to all consumers.[1]

An example will illustrate the pitfall in describing interaction frequencies as a means of inferring preferences amongst alternatives. If there are 125 people who all prefer large nearby towns to small nearby ones on the average three times out of four, and if all 125 are not so located to have a choice between these two alternatives, a description of the frequency with which each is chosen can result in a spurious inference about preference. Thus in Table 2.1 where only 20 have a choice between the two alternatives, out of 125 people 105 choose the small place, 20 the large one. If the researcher were to look at these choices <u>without regard to what alternatives are available</u>, the erroneous inference would be that the people prefer small nearby places to large nearby ones. This is precisely the kind of error that is made in models of consumer spatial preference using frequency of choice of alternatives as a basis. Clearly, if the analysis were repeated for

[1]The latter condition is only satisfied in the rare event that all consumers are located at the same place.

TABLE 2.1

Number of Households	Alternatives Available	Choice	
		large place	small place
5	Large nearby place only	5	-
20	Large and small nearby places	15	5
100	Small nearby place only	-	100
125		20	105

different sets of towns, a different conclusion about preference could probably be drawn from each analysis, provided the relative number of offerings of each alternative differed in each case. This is a common error in the behavioral analyses reviewed in this chapter.

The Regression Model of Preference

Most preference models seek to define preference as some function of environmental, and sometimes also, personal factors. Early attempts to derive consumer space preference functions sought to define a functional relationship between a consumer's choice of town for purchases and features of the town using a regression model. Thus for example Mitchell (1964) in a study of Iowa farm households sought to empirically define the parameters of the following functional equation:

$$T = a + bx_1 + cx_2 + dx_3, \qquad (2.1)$$

where T = the amount spent in a given town by an individual farm household,

x_1 = the distance between that farm and town,

x_2 = the population of the town,

x_3 = household characteristics, and

a, b, c and d = constants to be empirically defined.

In terms of the example in Table 2.1, equation 2.1 uses information on the number of households patronising places with small x_2 values and small x_1 values (105) and the number patronising places with large x_2 values and small x_1 values (20). Thus, the method completely ignores both the available, yet rejected alternatives, and the alternatives unavailable when one is chosen. Consequently the method has the same basic flaw as described above. To emphasise the inadequacy of the regression model where choices are made from different sets of spatial alternatives, Mitchell's percentage of explained variation of T using the above variables in only 35%. In fact Mitchell (1964, p. 85), without suggesting a better method, indicates that the above-mentioned inter-household variation in spatial opportunities is the cause of the low percentage of explanation.

> ...The different spatial positions of the various households with respect to the matrix of places of purchase around them, represents a rather complicated spatial pattern that regression-correlation is unable to hold statistically constant effectively.

The Gravity Model

Generally, the same criticism can be made of many gravity model formulations (not of the notion of the gravity model itself) as of the regression model. The explanation is that the gravity model has the same basic form as a regression model with two independent variables. Thus the basic gravity model

$$I_{ij} = k \frac{P_i P_j}{d_{ij}^{\alpha}} \qquad (2.2)$$

where P_i and P_j = the population at i and j,

d_{ij} = the distance from i to j,

I_{ij} = some measure of interaction between i and j, and

α and k = empirically derived constants,

can be rewritten when $P_i = 1$, as

$$I_{ij} = k \frac{P_j}{d_{ij}^{\alpha}} \tag{2.3}$$

Logarithmic transformation of equation 2.3 produces:

$$\log I_{ij} = \log k + \log P_j - \alpha \log d_{ij} \tag{2.4}$$

which has the form of a regression equation similar to equation 2.1.

Any preference model of this form where all individuals (i) are not similarly located with respect to spatial alternatives clearly may result in misleading inferences of the type already described. Many examples of such an application of the gravity model exist [Iklé (1954), Garrison (1956), Carrol and Bevis (1957), Marble (1959)].

However, one oft-quoted work by Huff (1962) whilst employing the above type of gravity model, is based on individuals whose location with respect to alternatives is virtually identical. The gravity model used closely predicts interactions between each of three neighborhoods and fourteen shopping centres (see Table 2.2). However, the high degree of accuracy is largely explained by the fact that between 70% and 90% of individuals in each neighborhood patronise the same centre (see Table 2.2) whose time distance is much less than any other centre (in neighborhood 1 it is 3.7 times as "near" as the next nearest centre, in 2 it is twice as near and in 3 it is 1.25 times as near) and whose square footage is exceeded by only 4 other centres. As a result the set of discrete points to which the regression plane is fitted (see Figure 2.1) has only one very large value and many very small values. To fit a regression plane to a small number of extreme frequencies and thereby obtain accurate results, is hardly surprising. If Huff had based the model on a set of frequencies covering also the intermediate ranges of time distance and centre size and obtained the same accurate fit, the results would indeed be interesting. However, as can be seen in Table 2.2 for neighborhoods 2 and 3, the few predictions where observed frequencies are not extremely high or low, have very high residuals indicating the poor fit of the regression plane to intermediate frequencies. Thus the apparent accuracy of Huff's model is largely a function of the over-simplified spatial choice situation on which it is based.

The Indifference Curve Model

In an attempt to overcome the weakness described in most regression and gravity models, Rushton (1966) uses a model which takes into account not only the frequency of choice of each spatial alternative but also the frequency with which each is available, though not necessarily patronised. The basic argument behind the preference model is that the higher the ratio of choice to availability of an alternative, the more preferred it is. Using data on the towns patronised for groceries by Iowa farm households (the same data as Mitchell's) this ratio was computed for many "town population/distance to town" combinations. Combinations with equal ratios were argued to be equally preferred and therefore the

TABLE 2.2

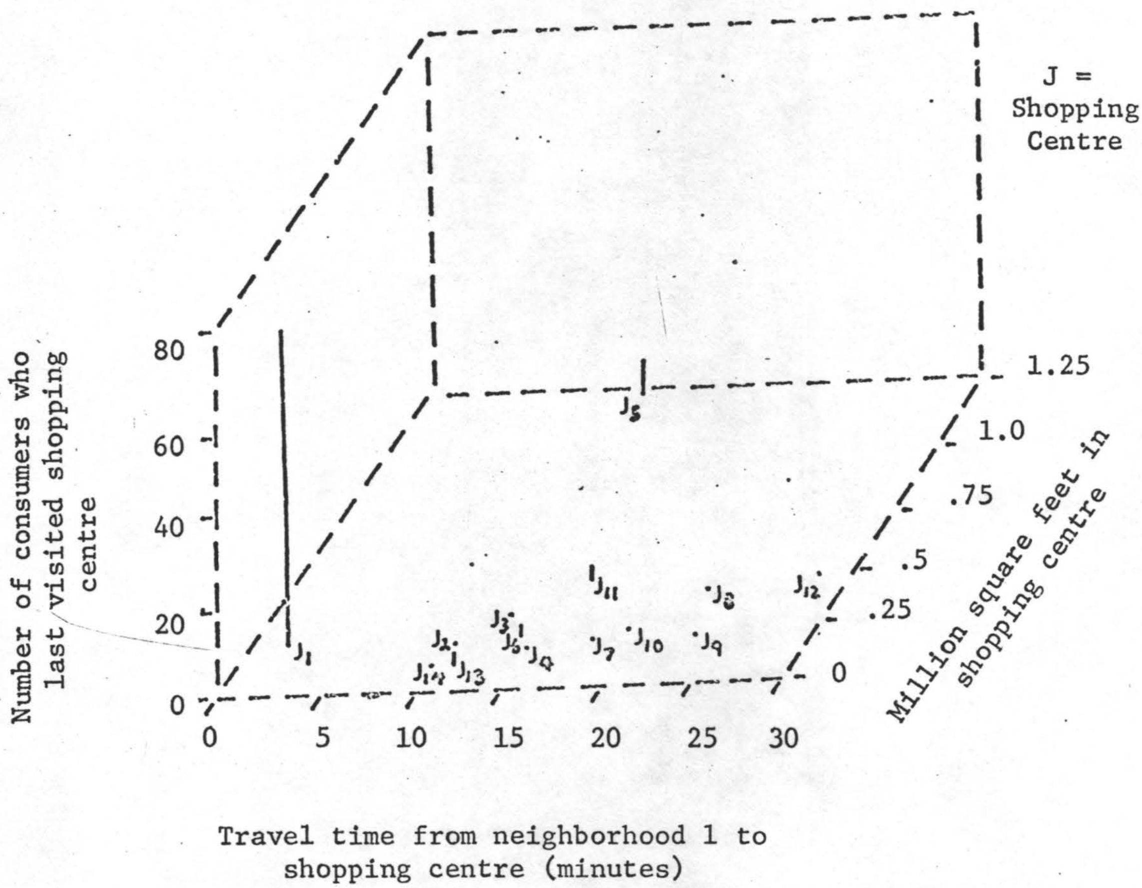
Comparison of Observed and Expected Number of Consumers from Each of the Three Neighborhoods Who Last Made a Clothing Purchase at One of the Specified Shopping Centres

Shopping Centre	Neighborhood 1		Neighborhood 2		Neighborhood 3	
	Observed	Expected	Observed	Expected	Observed	Expected
J_1	71	70.76	148	144.28	143	141.49
J_2	0	1.27	19	25.99	6	9.78
J_3	0	1.04	4	3.10	2	2.05
J_4	0	0.00	0	1.36	2	4.02
J_5	5	2.60	38	13.73	21	2.07
J_6	1	0.77	0	2.36	7	1.41
J_7	0	0.00	2	2.03	6	3.22
J_8	0	0.00	0	1.67	2	1.52
J_9	0	0.00	0	0.89	0	0.00
J_{10}	0	0.00	4	1.87	3	0.00
J_{11}	1	0.99	2	3.44	3	1.52
J_{12}	0	0.00	0	1.09	2	0.00
J_{13}	1	0.78	0	10.58	6	35.92
J_{14}	0	0.79	1	5.61	0	0.00
Total	79	79.00	218	218.00	203	203.00

Source: Huff, D. L., 1962, p. 454.

FIGURE 2.1[2]

The Spatial Interactions in Huff's Analysis by Consumers in Neighborhood 1 with the Selected Shopping Centres



[2] Figure derived from data in Huff's 1962 paper.

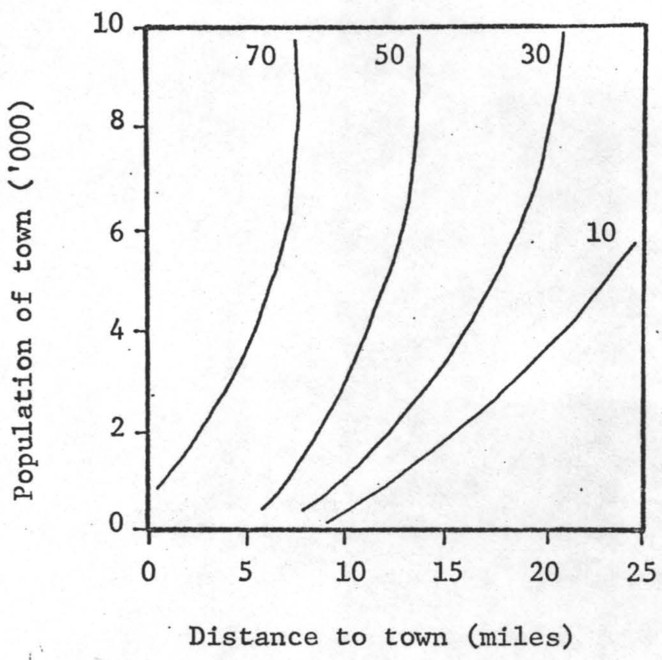
sample could be regarded as indifferent between such combinations. An indifference surface can be drawn indicating both the combinations between which there is indifference and the order of preference as revealed by the ratios (see Figure 2.2). Thus the combinations on the highest indifference curve are chosen in 70% of the cases they are available. Referring to Table 2.1 it is clear that this method of calculating preferredness is also unable to replicate the order of preference, as a result of ignoring what other alternatives are rejected when one is accepted. Thus, given the data in Table 2.1, large nearby towns are chosen 20 times out of the 25 times available (80%) whilst small nearby places are chosen with an even higher frequency, namely 100 times in 120 cases (83%). Once again as a result of the high frequency with which small places are available when large places are not, the erroneous inference would be that small places are preferred to large places, and not simply that they are chosen more often. As will be shown in an example in Chapter 3 using the same data, if only cases where both alternatives are available are used to calculate preference the correct preference order and proportion of preference are inferred. Rushton's indifference curve method, although not solving the problem, is the first spatial preference model to attempt to overcome the weakness of previous models which consider only the frequency with which alternatives are accepted.

Trade Area Studies

Although not specifically structured to elucidate empirical preference functions, trade area studies commonly assume, and sometimes claim to have proven, the Central Place Theory hypothesis of interaction

FIGURE 2.2

A Hypothetical Indifference Curve Model Showing
the General Features of Rushton's Model



with the nearest place offering the good desired. A trade area study conducted by Berry, Barnum and Tennant (1962) in southwestern Iowa demonstrates, according to Berry (1967, p. 16) that:

> ...the farmers make the same clear choice [in patronizing central places for goods and services]. There is only a little inter-digitation along the boundaries [of "trade areas"], and right along the edge farmers said they visited both centers, indicating that market area boundaries trace out real lines of indifference in choice.

No exact quantities are provided to reinforce the conclusion that households patronise the nearest place offering a good, although desire-line maps are produced.

A test of the accuracy of the conclusion is provided by Rushton in a study of a sample of the dispersed Iowa population (Rushton, 1966, p. 16). The hypothesis tested is that a consumer makes his maximum expenditure on groceries in the nearest town.[3] The definition of "nearest town" is successively varied to include only towns above a given population, with the results for each definition shown in Table 2.3.

TABLE 2.3

Results of Test of Nearest Town Hypothesis with Various Town-Size Groups

Hypothesis	Nearest town when only towns existing have population greater than:	Percentage of farm households correctly predicted
1	55	35
2	120	40
3	240	45
4	500	49
5	800	51
6	1200	52
7	2000	47
8	4000	37
9	7000	23
10	16000	11

Source: Rushton, 1966, p. 16.

[3] The trade area study is similarly "based upon responses as to where consumers purchased most of a given commodity" (Berry, 1967, p. 23).

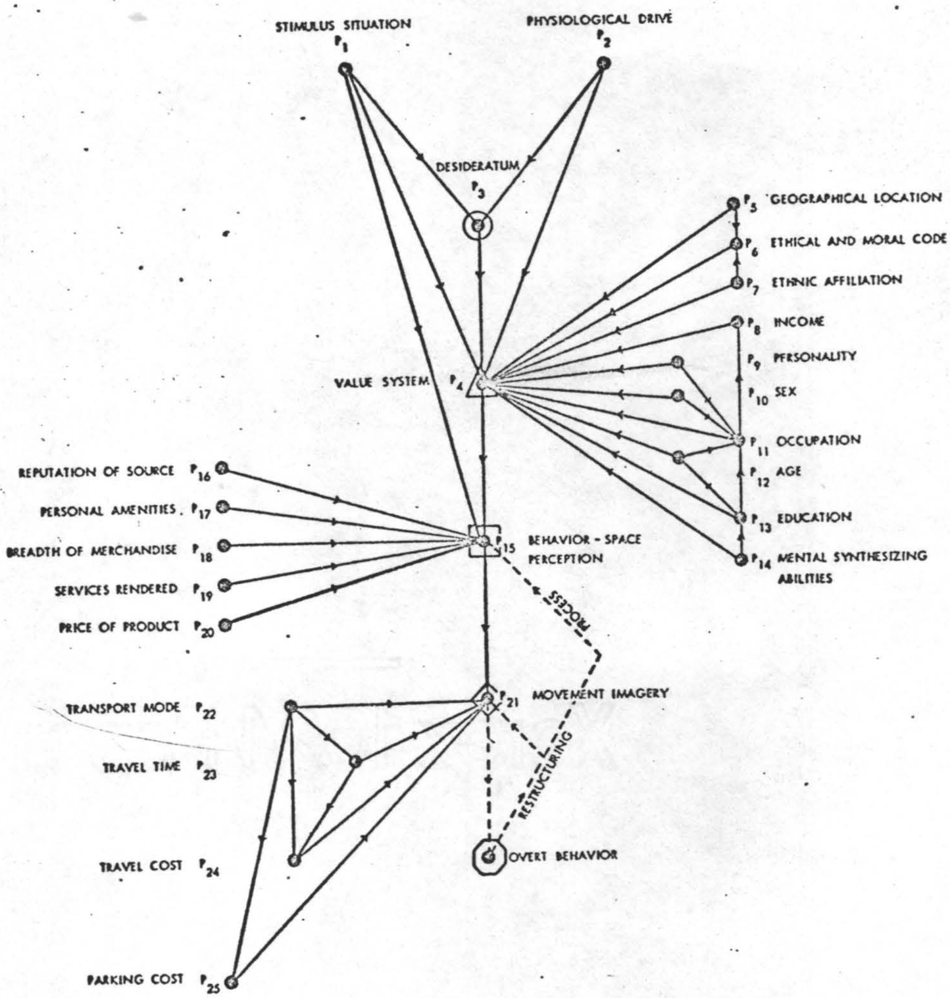
Clearly, no matter which definition of "nearest town" is used, no more than half the sample patronises the nearest town. This evidence, together with the lack of quantitative evidence in the 1962 study, and discrepancies between the evidence of some desire-line maps and stated findings, cast doubt on the validity and generality of the conclusion that consumers have a purely distance-minimising preference function.

Huff's Topographical Model

One other study merits a brief commentary, if only to illustrate the inappropriateness of its title, namely Huff's "topographical model of consumer space preferences" (Huff, 1960). In his paper he develops a *sui generis* model incorporating several groups of spatial and non-spatial variables, intuitively deduced to be related to consumer space preference functions (see Figure 2.3). Those variables described under the headings "behavior-space perception" and "movement imagery" can be defined as "attributes of spatial alternatives", whilst those related to the consumer's "value system" describe "attributes of the decision-maker". Although Huff attempts to deduce the connectivity between all variables in the model, the study cannot utilize intuitive deductive methods to evaluate the connectivity between each variable and preference, and more fundamentally, the relative degree of connectivity between attributes of spatial alternatives and consumers and their preference as revealed through overt behavior. Huff's study simply describes the *hypothesized* relationship between factors which in turn he can only hypothesize to be related to consumer spatial preference. By contrast, the other "spatial" studies discussed in this chapter do attempt to demonstrate the relationship between attributes of the environment and spatial preference. Thus what he describes as a "model of consumer space preferences" is simply a discussion of factors which may or may not be related to consumer spatial preference.

FIGURE 2.3

Basic Interactions of Huff's Topographical Model



Source: Huff, D. L., 1960, p. 165.

A Model of Residential Preference

An interesting attempt is made by Peterson (1967a, 1967b) to determine the role of certain attributes in contributing to individuals' "preferences" for residential neighborhoods. The model is based on the scores assigned by 140 individuals to 10 attributes in 23 photographs of residential neighborhoods in suburban Chicago. The following attributes had to be scored:

Preference
Greenery
Open Space
Age
Expensiveness
Safety
Privacy
Beauty
Closeness to Nature
Quality of the Photography

A regression analysis[4] with "preference" as the dependent variable, and the other 9 variables as independent variables results in 79% of preference variation being explained by the following equation:

$$\text{Preference} = -0.56\ (\text{Age}) + 0.46\ (\text{Closeness to Nature}) \qquad (2.5)$$

Thus the inference might be made that preference for residential neighborhoods is largely an inverse function of their age and secondarily of their closeness to nature. However several cautions are warranted. Firstly, the entire study is based on people's introspective evaluation of their sentiments regarding the visual appearance of a place, which may differ

[4] This experimental analysis in contrast to most spatial behavior analyses is one case in which regression analysis is a valid technique, since all individuals are presented with the same set of alternatives to which to assign scores.

from their actual subjective evaluation of the appearance. Thus as Peterson suggests (1967a, p. 30), the "conclusions...are presented only as refined hypotheses, because they are based on observed correlations among psychologically obtrusive measurements". Secondly, the preference measured is one for visual appearance which may differ from preference regarding residential desirability. Thirdly, the sample of individuals "was not obtained by any rigorously random process, [and therefore] no external validity can be claimed" (Peterson, 1967a, p. 22).

Thus whilst Peterson's analysis does not contribute to an understanding of preference as revealed through behavior, it does provide an example of the potential in experimental spatial preference analysis for hypothesis generation.

Rushton's Scaling of Locational Preference

Appreciation of the basic flaw of "directly describing the behavior occurring within a...system" (Curry, 1967, p. 219) as a means of inferring spatial preference is largely attributable to Rushton. He succeeds in overcoming the flaw by using the psychologists' Method of Paired Comparisons on spatial choice data as a means of reconstructing preference structures (Rushton, 1969a, 1969b). However, since the present study is based on the same method, but includes several extensions of and improvements on Rushton's work, a discussion of the method and criticism of parts of Rushton's work are left to Chapters 3 and 4. Suffice to say, it will be demonstrated here that the Method of Paired Comparisons overcomes the basic flaw discussed with regard to regression models and most gravity model formulations and can therefore be used to accurately reconstruct a sample's aggregate preference structure within the limits set by sample size, scale of analysis, and the definition of spatial alternatives.

Models of Space Preference Differences

Several studies in the past decade have sought to distinguish differences in space preferences related to different types of individual, different time periods or different trip purposes. The earliest empirical test is by Huff (1962), who sought to compare the α coefficients, such as occur in equations 2.3 and 2.4, which are empirically derived for each of three neighborhoods using least-squares regression. Each α value is, however, derived from a regression analysis based on different sets of spatial alternatives, in that each neighborhood is differently located with respect to the shopping centres. Thus the differing interaction frequencies with different "distance/size" shopping centres, which result in regression planes with differing slopes and hence differing α values, may be a function of the differing spatial opportunities of each neighborhood. Thus Huff avoids the error of a regression analysis of differently located consumers, by separately analysing each neighborhood where consumers are similarly located, but makes the error by comparing regression results for differently located consumers. Empirical evidence that the difference in interaction frequencies of the three neighborhood (see Table 2.2) is probably due to their different locations, is that neighborhood 1, which has the most extremely biased patronage frequency, is also the neighborhood which has the most biased location, in terms of being closest to the most patronised centre (J_1) and furthest in the aggregate from all others. For example J_1 is 2.8 minutes distant while the next nearest is 10.4 minutes distant, i.e., a ratio of 3.7 times, whereas for the other two neighborhoods the ratio is only 1.9 and 1.25.

Nevertheless, a comparison of α values obtained separately in each neighborhood for clothing purchases and furniture purchases is methodologically reliable insofar as the α value for each commodity can be compared for the same neighborhood. The result is, as should be expected, that the α value is consistently lower for the higher order good, furniture, by proportions of 10%, 12% and 19%. Thus the evidence is, as expected, that consumers are willing to travel farther for furniture than for clothing.

Mayfield (1963) purports to provide evidence of differing space preferences in a study of trips by Punjabi villagers to towns to purchase new milled white cotton cloth. He notes a high positive correlation between the number of central functions at the centre visited and the "round-trip distance to the nearest alternate central place of a functional size $\geqslant$ central place where good purchased" (Mayfield, 1963, p. 46). On this basis he reaches the following conclusion:

> ...the consumer has selected a place in a particular functional size class for his purchases. Thus a preference for a certain level of interaction has contributed to the consumer's concept of the importance of a central place. (1963, p. 47)

In comment, it seems that the correlation discussed reflects a basic property of central place systems in general, i.e., that the larger and more functionally diverse the town, the greater is its distance to another town of equal or greater size and diversity. It is inconceivable to this writer how "a preference for a certain level of interaction" contributing "to the consumer's concept of the importance of a central place" can be inferred from the above correlation.

On the basis of a cluster analysis, "population of the village of trip origin" and "population of that central place where the good is purchased" are grouped in the same cluster. Insofar as the initial correlation matrix is not shown, it is not known whether the correlation between village population and the patronised centre's population is positive or negative. On the basis of this correlation the following conclusion is reached:

> The next factor to be added is "population of the village of trip origin". It was noted that within the social structure of Punjab society, an individual in a large village usually had a greater number of social contacts possible. Such contacts, in adding to the level of information of the individual, might encourage him to raise the level of his desire for spatial interaction. That is, a consumer in a large village might give more importance to a large center than would a consumer from a smaller community. This, then, is also a factor in the level of importance of a central place, as indicated by space preferences. (1963, p. 47)

Two cautions are required. The first is that no indication is provided as to the size or direction of correlation, although the latter is presumably positive. Secondly, no test is performed on the central place system in the study area to ensure that there is no local tendency for larger villages to be closer to larger towns and smaller villages to be closer to smaller towns. Such a situation might arise, for example, as a result of local variations in crop carrying capacity, so that some areas could support larger villages and towns than others. Thus of the two conclusions regarding differing space preferences, the first seems entirely spurious and the validity of the second cannot be checked.

More recently Murdie (1965) compared the spatial expenditure patterns of Old Order Mennonites and "modern" Canadians in Waterloo County, Ontario. The specific behavioral characteristic measured was "distance travelled to the first choice center" where "first choice center is the place where a good is most frequently obtained" (Murdie, 1965, p. 215). Defining that distance as the dependent variable in a regression equation and the number of central place functions as the independent variable, Murdie tests if the regression coefficients are significantly different between the two groups for each of eight goods and services. Whilst visual inspection of distribution maps of the 95 sample farm households in each group suggests there is no great difference in the location of each with regard to urban centres, Murdie provides no statistical proof of this. Thus significant differences in the regression coefficient could indicate the groups are different distances from the same retail outlets and not necessarily, as Murdie suggests that their space preferences differ. Assuming, however, that the discrepancy in the two groups' distributions is minimal, Murdie finds significantly differing preferences regarding the following purchase trips:

shoes,
clothing (and yard goods),
food, and
auto repair-harness repair.

No significant differences are observed in trips for:

doctor,
dentist,
bank, and
appliances.

Murdie explains preference differences in the former group in terms of their being "traditional" functions "for which patterned behavior existed when the Old Order Mennonite culture was first established in this area". By contrast the other activities are higher order activities and necessitate both groups normally visiting a place other than the nearest village. Thus with some reservations about the absence of a test for similarity in the two groups' distributions, Murdie provides an example of the marked spatial preferences differences which can occur when two groups have very different cultural mores affecting their mode of transport.

A study by Ray (1967), also in Ontario, tackled this same notion of differing space preference with respect to anglophone and francophone communities in the same area of eastern Ontario. Unlike Murdie's study, the two groups are located almost exclusively in two different areas adjacent to one another. In the following abstruse passage, Ray (1967, p. 151) seems to suggest a method for testing for different space preferences for differently located groups:

> Two hypotheses are tested for each pair of groups. The first hypothesis tests differences in the covariance structure of trip length, and ignores the mean distances travelled, which reflect the geographic location of the farm families in relation to the service centers. The second hypothesis is that no significant differences occur in the means and covariance structure of the distances travelled for services by the two groups being tested. Any two groups are considered to have different travel behavior only where both tests are significant.

Several points are unclear. Firstly, it is not explained how a covariance analysis can be performed on groups whose members are not paired according to any rule. Secondly, allowing for a covariance analysis of random pairs for each pair of groups, it is unclear how significant differences in the mean and variation of trip lengths for two groups allows the inference to be drawn that travel differences are more than a function of differing spatial opportunities. In light of the lack of explanation of the tests provided by Ray, it is difficult to conclude whether the behavioral differences observed for the two groups are a function of location, or of preference, or of both.

Rushton (1966) performed a similar test for differing space preferences in different social and economic groups of Iowan farm and non-farm households. Having sub-divided the sample population according to a given socioeconomic variable, the null hypothesis tested was that there was no significant difference in a given aspect of the spatial behavior of the groups (e.g., distance travelled to the maximum grocery purchase town). The t test was used to determine if the group means were significantly different, and the F test to detect whether between-group variance was significantly greater than within-group variance. However, no analysis was performed on each social and economic grouping to test whether for each household in one group there was one in the other with the same or very similar set of spatial opportunities. Thus, without this evidence, it is impossible to determine whether the significant between-group differences in behavior which were established in the F tests and t tests, are a function of spatial situation, or of differing preferences, or of both.

A similar criticism can be levelled against a comprehensive analysis by Bucklin (1967) of shopping behavior in the Oakland, California area. Initially he uses a regression equation to describe the proportion of shoppers in a census tract interacting with a shopping centre as a function of distance between tract and centre. Since shoppers in each tract are differently located with respect to shopping centres, the use of regression is subject to the same error as the regression studies discussed earlier in the chapter.

A less obvious example of the misuse of regression in determining preference functions is found when he seeks to determine the variables explaining different shoppers' patronage of different centres. In this case shoppers are grouped according to the centre patronised and with information on 55 characteristics of each shopper, linear discriminant functions (LDF's) are derived to explain shoppers' different choices as a function of their variable scores. An LDF takes the form of a multiple regression equation without a constant, i.e.,

$$Y = b_1X_1 + b_2X_2 + \ldots\ldots b_pX_p \qquad (2.5)$$

where X_p = a shopper's score on the pth variable,

$b_1, b_2, \ldots, b_p$ = empirically derived coefficients, and

Y = a dichotomous variable in the case of a two-way discriminant function.

Thus in the simplest case of a two-way discriminant analysis, the LDF is derived which provides the maximum difference between the mean discriminant score ($\overline{Y}$) for the two groups, where $\overline{Y}$ is computed using equation 2.5 with the $\overline{X}$ values of each group substituted for the X values of each shopper. For each shopper a discriminant score (Y) is calculated using equation 2.5 and the shopper assigned to one of the groups according to which mean discriminant score ($\overline{Y}$), its score is closer to. Bucklin uses this discriminant information to calculate the probability of a shopper with given characteristics patronising one of several centres. For example, using LDF's, blacks are more commonly assigned to the group patronising downtown Oakland than to groups patronising less central shopping centres. However, such a propensity for downtown shopping may simply indicate blacks live closer to downtown, despite Bucklin's unsubstantiated assertion that this bias in patronage indicates "an attraction beyond the simple fact of Negroes' residential proximity to this area" (Bucklin, 1967, p. 85). Thus the greater probability of a certain type of shopper patronising a certain centre may be due to opportunistic location as much as to preference. Linear discriminant functions per se, like regression functions, are not designed to eliminate the effect of spatial bias in shoppers' spatial alternatives. Hence the conclusions drawn by Bucklin about the effects of demographic, motivational and other attributes on spatial preference are subject to the same possible error as the regression and gravity analyses described earlier.

A more recent work by Rushton (1969b) overcomes the problem of the differing sets of spatial alternatives of members of different groups. Using the Method of Paired Comparisons, which is described in detail in Chapters 3 and 4, he compared the preferences of two samples of rural Iowa households, one taken in 1934, the other in 1960. Rushton concludes that a decrease in distance minimising objectives resulting in a greater patronage in 1960 of larger more distant towns has caused small towns and villages to lose business and either become less viable or disappear altogether. However, certain weaknesses in the method of comparison should be noted. Firstly, the scale of analysis is different for the two groups, since thirty classes of alternative opportunities are defined for the 1960 sample and forty-two for the 1934 sample, although the entire domain of alternative opportunities is the same for both groups. Thus a finer scale of analysis is used for the 1934 sample, and therefore apparent preference differences could be wholly or partly a function of the different scales of analysis. A second weakness is that no allowance is made for the sampling error associated with each preference structure, in that there is no check as to what differences might occur between the preference structures of two random samples drawn from the same population. Thus the observed difference in preferences between the two samples might be due in part or in whole to sampling error. Thirdly, the preference surfaces, which are drawn in the manner of the indifference surface shown in Figure 2.2, are linearly interpolated between scale values which are ascribed to the mid-points of alternative opportunities having a form similar to those in Figure 3.2. It is possible, however, that the mean of spatial opportunities actually available to each group in an opportunity class differs from the mid-point. Hence the basis of the interpolation may be inaccurate. Finally the comparison of the two surfaces drawn is based on nothing more rigorous than visual inspection. There is no way of knowing whether the above potential sources of error are cumulative or cancel out in practice. Without evidence as to the strength of these sources of error, it is impossible to say how accurate are the intuitively appealing preference differences which Rushton claims to demonstrate.[5] Notwithstanding the above caveats, the study is probably the soundest analysis of differing space preferences to date.

One other group of studies in the field of retail consumer space preferences deserve mention, if only because their conclusions are often referred to in the literature as showing that different types of consumer have different space preferences. These are studies of consumer trip frequencies. Marble (1959) in a comprehensive review of this literature finds that several authors have been able to relate trip frequency to the social and economic characteristics of individuals and households (Gardner, 1949; Hamburg, 1957; Mertz and Hamner, 1957; Bureau of Population and Economic Research, 1951; Jonassen, 1955). The findings of Marble (1959) in his Cedar Rapids, Iowa study are often improperly cited as empirical evidence of differing space preferences. Yet as Marble suggests (1967, p. 39) the study only:

> points out the importance of the socioeconomic structure of the household in determining such things as gross trip frequencies and total time spent away from the home.

In other words, the relationship found is with temporal aspects of behavior and not spatial and therefore no conclusion can be drawn as to whether differences in trip frequencies indicate differences in spatial preferences.

A study by the sociologist Michelson (1966) focuses attention on the relationship between personal attributes and the ideal urban residential environment. The conclusions drawn are based on "lengthy

---

[5]The present analysis overcomes all the weaknesses outlined above.

interviews with a sample of 75 respondents" (Michelson, 1966, p. 357) followed by content analysis of the interview transcripts. Individuals were shown photographs of buildings and neighborhoods representing variations within each of a number of dimensions of environment. The individual was amongst other things, required to rank the photographs and give associated "value rationales". In using photographs, the criticism can be made that the subject may be indicating preference for appearances rather than residential preference. Likewise the study is based on people's introspective estimation of their preferences, not on preference as revealed in real-world choice situations. His conclusions are as follows:

1. The social variable explaining most variance in the amount of separation people choose from others' homes is the distance people now live from their personal friends. This result, in a group where every family has a car, might simply indicate nobody is inconvenienced by the distance they live from friends. Or they may prefer to give the interviewer an impression of satisfaction with personal relationships.
2. Rather axiomatically, people who value convenience, prefer objects with which they interact spatially, to be near at hand.
3. "People who are relatively cosmopolitan in their choice of stores, who do not patronise an establishment merely because it is the closest one, are more apt to desire objects perceptually separated from their homes." (Michelson, 1966, p. 357) Again this seems somewhat axiomatic.
4. "The scale of objects desired varies inversely with expressions of

class consciousness. People who want small stores, churches, and the like tend generally to speak of economic and social inequalities and their personal desire to limit interaction to their own class". (Michelson, 1966, p. 357). Once again such a relationship seems axiomatic. It bespeaks the exclusiveness of the rich "wasp" neighborhood as much as the introversion of the ethnic neighborhood.

5. The popularity of the single family house is so widespread that its choice is independent of any social variable analyzed.
6. "Two major types of variable are conspicuous by their failure to vary systematically with ideal choices of environment. One is social rank, and the other is stage in the life cycle.....The choices people would make are not a simple function of their age or status, but of more subtle influences - their values and styles of life." (Michelson, 1966, p. 358).

In his first four conclusions, Michelson seems to have shown, not surprisingly, that some social values or preferences result in certain urban environmental preferences. His fifth conclusion indicates that for many different types of individual the single family house lies at the top of the "residence type" preference scale and has no close competitors. His sixth conclusion, if reliable, is a useful debunking of commonly held beliefs that status and stage in life cycle affect residential preference. What is undoubtedly true is that status and stage in life cycle affect the individual's socially or economically feasible range of alternatives for such things as residence type. Michelson's study is concerned with preference amongst complete ranges of alternatives whether or not it is feasible for the individual to attain his ideal.

On the basis of conclusions 2, 4 and 6, one hypothesis of use to the geographer which might be formulated,is that the personal attributes most likely to be predictors of differing spatial preferences are other of the individual's more basic preferences, spatial or otherwise, rather than social or economic variables.

Conclusion

The foregoing review has sought to demonstrate that until recently methodological weaknesses in space preference analyses have prevented almost any reliable conclusions as to the nature of consumer space preference functions. The studies by Peterson and Michelson which avoid these weaknesses have been concerned, however, with preference as revealed by introspection rather than the more reliable source of overt real-world choice. The only studies with a sound basic methodology which have been concerned with preference revealed through overt choice are the two most recent ones by Rushton (1969a, 1969b). As mentioned already, there are some weaknesses in both these studies. The present study seeks to overcome these problems and in addition to perform further analyses of consumer space preference functions, which are described in Chapter 3.

Chapter 3 THE RESEARCH DESIGN

In Chapter 2 it was established that in situations where all alternatives are not available to everyone, a model of preference which considers only chosen alternatives and not those rejected, is likely to provide spurious descriptions of preference. In addition, a basic premise is that the most reliable source of preference information is not what people say they prefer, but what their choices reveal them to prefer. Thus any improved preference model should be based on the axiom that the preferredness of any one thing over another is discernible in behavior only if both are available and one is chosen. What is required, therefore, in order to reconstruct the preference structures on which choices are based, is a method of handling choice data where choice is considered in the context of the alternatives available rather than in the vacuum that considers only alternatives accepted and not those rejected. Only then is it possible to describe rules of spatial preference which are independent of any particular arrangement of alternatives. Such descriptions would avoid the spuriousness of previous models and, in so doing, would have the added distinction of being more likely to be reliable statements of preference outside the particular study area used in the original analysis.

The Method of Paired Comparisons

One appropriate technique for handling choice data in this manner is the Method of Paired Comparisons developed in psychology. It is a means of inducing individuals' preference structures from their choices between pairs of alternatives, by presenting them with all possible pairs of alternatives. To obtain reliable information on which of a pair is more preferred, a large number of comparisons are required. Such comparisons may either be made repeatedly by the same individual or singly by many different individuals, with the more preferred member of a pair being that which is chosen proportionately more often. Thus in the hypothetical example described in Chapter 2, where people preferred large nearby towns to small ones three out of four times on average, the task, using the method of paired comparisons, would be to reconstruct this preference, using as data only household choices between the two types of town. For the 20 out of 125 households so located as to have this choice, 15 chose the larger town, from which the correct inference would be drawn that large nearby towns are preferred to small nearby ones three times out of four.[1] This is the simplest example of how the method of paired comparisons avoids the logical pitfall, encountered in other geographical analyses, where people's preferences are inferred from their frequency of choosing each alternative, regardless of what alternatives are simultaneously available.

The method's accuracy in deriving preference from choices in a more complex situation is shown in Figure 3.1.

---

[1]Clearly, no preference information is available from the other 105 households who did not have both types of alternative available.

FIGURE 3.1

(a)

RANK ORDER OF PATRONAGE OF ALTERNATIVES

ALTERNATIVES

HOUSEHOLDS	A	B	C	D
1	1	2	X	
2		1		X
3			1	X
4		1	2	X
5	1			X
6			1	X
7			1	X
8		1	2	X
9			1	X
10	1	2		X
11			1	X

1 = first preference alternative

2 = second preference alternative

X = alternative available but not patronised

blank = alternative not available

⟹

(b)

AGGREGATE PROPORTIONATE PREFERENCE OF ROW ALTERNATIVE OVER COLUMN ALTERNATIVE

	A	B	C	D
A	/	1.0	1.0	1.0
B	0.0	/	1.0	1.0
C	0.0	0.0	/	1.0
D	0.0	0.0	0.0	/

⇓

A > B > C > D

In this hypothetical example, all households are posited to have the same preference ranking of alternatives, namely A > B > C > D, where > symbolises "is preferred to". In addition it is a consistent preference rule such that each alternative is either absolutely preferred to, or absolutely preferred by each of the other alternatives. In Figure 3.1(a) the choices or rejections of alternatives are indicated. It is assumed that in a choice set where more than two alternatives are available, implicit paired comparison information exists for all pairs of available alternatives where one of the pair is chosen. Thus if an individual chooses town A when towns A, B, C, D are available, he is assumed to reveal implicit paired comparison information of the form A > B, A > C, A > D. Note, however, that his choice provides no information on his paired comparisons of B, C, and D, since he does not select any of these in preference to the others. Likewise, of course, there is no information on his paired comparison between A and any other unavailable towns. In Figure 3.1(a) household 1, then, is regarded as revealing pair-wise preferences between A and B, A and C, and B and C. The sum total of the pair-wise preferences of the 11 households is shown in Figure 3.1(b), where A is seen to be preferred to all other alternatives, B is preferred to all others except A, and so on to D which is preferred to no others. Thus the method permits the accurate reconstruction of the preference structure upon which all the choices are based, despite differences in the alternatives available to each household. By contrast if the preference structure were induced from measures of the frequency of choosing each alternative, the erroneous conclusion would be that C > B > A > D. Likewise, if the structure were inferred from the frequency of each alternative's choice as a first preference alternative, the inference would be that C > B, that there is indifference between B and A, and that A > D. Thus the method of paired comparisons would appear to be a more reliable methodological basis for the analysis of preference than methods previously used in geography.

The Data

The data used in this analysis of consumer space preferences are of the type indicated in Figure 3.1(a) where each household chooses some towns for shopping purposes from a given subset of alternatives. These subsets of alternatives normally vary from one household to another. Specifically the data describe the location of households in a stratified areal random sample of Iowa farm households (530 households), together with the towns where each bought groceries in 1960, and the grocery dollar expenditure in each town that year. In addition the population and location of the 1132 towns in Iowa with a 1960 population over 50 is given, these being treated as the entire set of alternatives available to each household.[2,3] Thus for each household there is information on which towns it chose for shopping, its ordering of these in terms of dollar expenditure, and the many towns which were not patronised. This provides an adequate basis for paired comparison information of the type described above.

---

[2] As a result, information on households which crossed the state boundary for grocery purchases is not used. Since this amounts to a very small number of information bits, it is unlikely to affect the analysis.

[3] A more complete description of the sample and sample design is provided in Appendix A.

The Definition of Spatial Alternatives

As presently described, each household chooses from amongst a unique set of spatial alternatives, insofar as no two households in the sample are identically located and therefore no two households have the same spatial relationship to all towns in Iowa. A hypothesis tested in this study, however, is that towns have varying degrees of attractiveness for shopping patronage that are related to specific attributes of the town. In other words, towns can be multivariately defined in a "town attribute" space, such that the attractiveness of a town is a function of its position in that attribute space. In the same sense that in Chapter 1 the attractiveness of different houses was described as a function of their neighborhood characteristics, construction quality, etc., so can alternative towns be hypothesized to vary in attractiveness for shopping as a function of those attributes considered relevant by consumers. It is possible to conceive of many such attributes as relevant to spatial choice, including the diversity of retail and service functions, the number of retail outlets for each function, product price, the reputation of the outlet, travel time between household or workplace and outlet, parking and travel costs, and utilities of the place not directly related to consumption such as friendship or kinship ties. There are two reasons for not including a large number of town attributes in the paired comparison model. Firstly, some of the above attributes are either strongly correlated with one another or with some other attribute. Thus for example it has been shown in south-west Iowa that the population of a town is strongly correlated with both the number of functional units (r = .98) and the number of central functions (r = .89) (Berry, Barnum and Tennant, 1962). Similar evidence is provided by Stafford (1963) and Thomas (1960). In addition, urban congestion and therefore intra-urban travel time is often correlated with a town's population. In a state such as Iowa, with a relatively even terrain and a network of relatively straight roads with little variation in possible driving speeds, it is also reasonable to assume a strong correlation between time distance and road distance between consumer and town. In view of such correlations between attributes it is possible therefore to collapse the previously-mentioned multi-attribute space into as few as two relatively independent dimensions relating to the town's size or population and its distance to the consumer.

The second constraint on inclusion of many attributes in the model is that for every addition of another attribute, the increase in the total possible number of pairs of alternatives, defined in terms of these attributes, increases in a geometric rather than arithmetic progression. Consequently, with a sample of only 530 households many pairs of alternatives would likely never be compared, or at best so rarely as to prevent any accurate inference of preference.

Furthermore, considerable evidence in the literature points to the above-mentioned attributes of town population and distance or their correlates as being major factors affecting consumer spatial choice of towns for retail patronage. Specifically, this relationship is demonstrated empirically in works by Berry (1967, pp. 10-20), Huff (1962), Mayfield (1963), Rushton (1964, 1966), and Thomas, Mitchell and Blome (1962). Thus it is hypothesized here that town population and distance to the consumer (or their surrogates) are two critical attributes of a town's utility as a shopping place. Specifically, it is hypothesized that in general its utility is directly proportional to town size and its correlates, although there may be a point in the town size continuum at which marginal utility becomes negative. This refers to the situation where the traffic congestion commonly associated with larger cities results in such towns being less attractive than somewhat smaller ones.[4] It is further hypothesized that there is an inverse relationship between the utility of a town and its distance from the consumer.

Whilst the outcome of the test of the above two hypotheses is hardly in doubt, there is doubt about the extent of their contribution to an explanation of variation in consumer spatial choice. A test of the extent of their contribution is described in Chapter 4. The need to add other independent attributes of spatial alternatives to the preference model depends, of course, on the extent of unexplained variation in the present two-variable model. Therefore, the question of how satisfactory are the attributes chosen, is left for consideration in Chapter 4.

However, with regard to the argument that two attributes are perhaps too few to include in a preference model involving complex choice situations, interesting conclusions have been reached in psychology about how many attributes or conceptual units the individual can weigh, quantitatively perceive and combine in any choice situation.

---

[4]See Rushton (1966, p. 37) for tentative evidence supporting this conclusion.

According to Shepard (1964, p. 263) the evidence indicates that:

> there are rather severe limitations on the number of conceptual units that can be handled at any one time. Moreover, although a small number of attributes evidently can be combined according to simple linear or additive rules,.....non-linear rules or complex interactions between variables seem to offer great conceptual difficulty.

Hence the use of only two attributes to define spatial alternatives seems acceptable.

An Operational Definition of Spatial Alternatives

The definition of a spatial alternative as any point in a continuous bivariate space, whose dimensions are town size and distance to consumer, is inadequate for the present paired comparison analysis. The explanation is that there would be few, if any, cases of the same two points being compared more than once. For rarely, if ever, in this sample, will more than one household choose from precisely similar pairs of alternatives in terms of town population and distance. With a sample size of one associated with each paired comparison, the sample error of each preference estimate would be extremely large and as a result little useful inference could be drawn. Thus whilst a punctiform definition of spatial alternatives would permit analysis that was not scale-dependent, it would be virtually impossible to operationalise. For pragmatic reasons, therefore, it is necessary to group different town population/distance combinations as the same spatial alternative. However, there does appear to be behavioral justification for such an aggregation of "different" spatial alternatives, which for the purposes of analysis means that all town population/distance combinations within given bounds are treated as the same spatial alternative. The notion of imperfect discrimination, mentioned in Chapter 1, implies in effect that there are limits to the individual's resolving power, so that things which are similar but not identical nevertheless are perceived to be the same. Thus at a certain distance, two letters such as O and D may be visually indistinguishable. Similarly if two towns have similar amounts of the attributes relevant to choice, the consumer may not perceive any significant difference between them and hence be indifferent as to which he chooses.

As early as the mid 19th century, psychologists investigated how large the difference in the amounts of an attribute possessed by two alternatives must be for the individual to perceive a difference (Fechner, 1966; Brown, 1910). Usually though, psychologists have attempted to define that amount of difference between two things such that on only half of the trials, subjects notice the difference. This amount of difference is described as a "just noticeable difference" (jnd for short). Whilst the concern in this study is for the amount of difference which is never or hardly ever noticed, rather than that noticed half of the time, conclusions reached regarding the jnd would seem to be equally applicable to perceptions of no noticeable difference and the related preference concept of indifference between "like" objects. The major conclusion of relevance is that the jnd is a fixed proportion of the size of the "standard stimulus". Thus if the standard stimulus is for example a 20 lbs weight, the jnd may be 2 lbs, whereas, if the standard is 40 lbs, it will be 4 lbs. Translating this into attributes relevant to consumer town choice, it may be that a consumer who barely distinguishes between a town of 200 and 300 is unlikely to distinguish between one of 2,000 and 2,100 even although their absolute differences are the same. If the jnd is a fixed proportion, for example 30%, then the difference between towns of 1,000 and 1,300 will be distinguished as often as that between towns of 10,000 and 13,000. Thus, in defining spatial alternatives by town size, it would be inappropriate to have equal-sized town population categories, but rather categories whose sizes increase in proportion to the population, e.g., 1,000, 3,000, 9,000, 27,000, etc. The same argument applies to distances. The difference between 3 and 6 miles is more likely to be perceived than that between 23 and 26 miles. Thus it is appropriate to apply the same rule to distance as to town population. In addition, however, it seems intuitively obvious that whilst a town of 200 at 3 miles may be perceived as different from one of 200 at 5 miles, a town of 20,000 at 3 miles may appear similar to one of that size at 5 miles. Thus it seems that the "distance to town" limits used in defining similar spatial alternatives should vary not only as a direct function of the distance but also as a direct function of the town population. Thus for example, appropriate distance limits for alternatives with populations less than 200 might be 0 to 2 miles, 3 to 6, and 7 to 12, whilst for towns between 3,000 and 10,000 more appropriate limits would be 0 to 5 miles, 6 to 13, 14 to 22, etc.

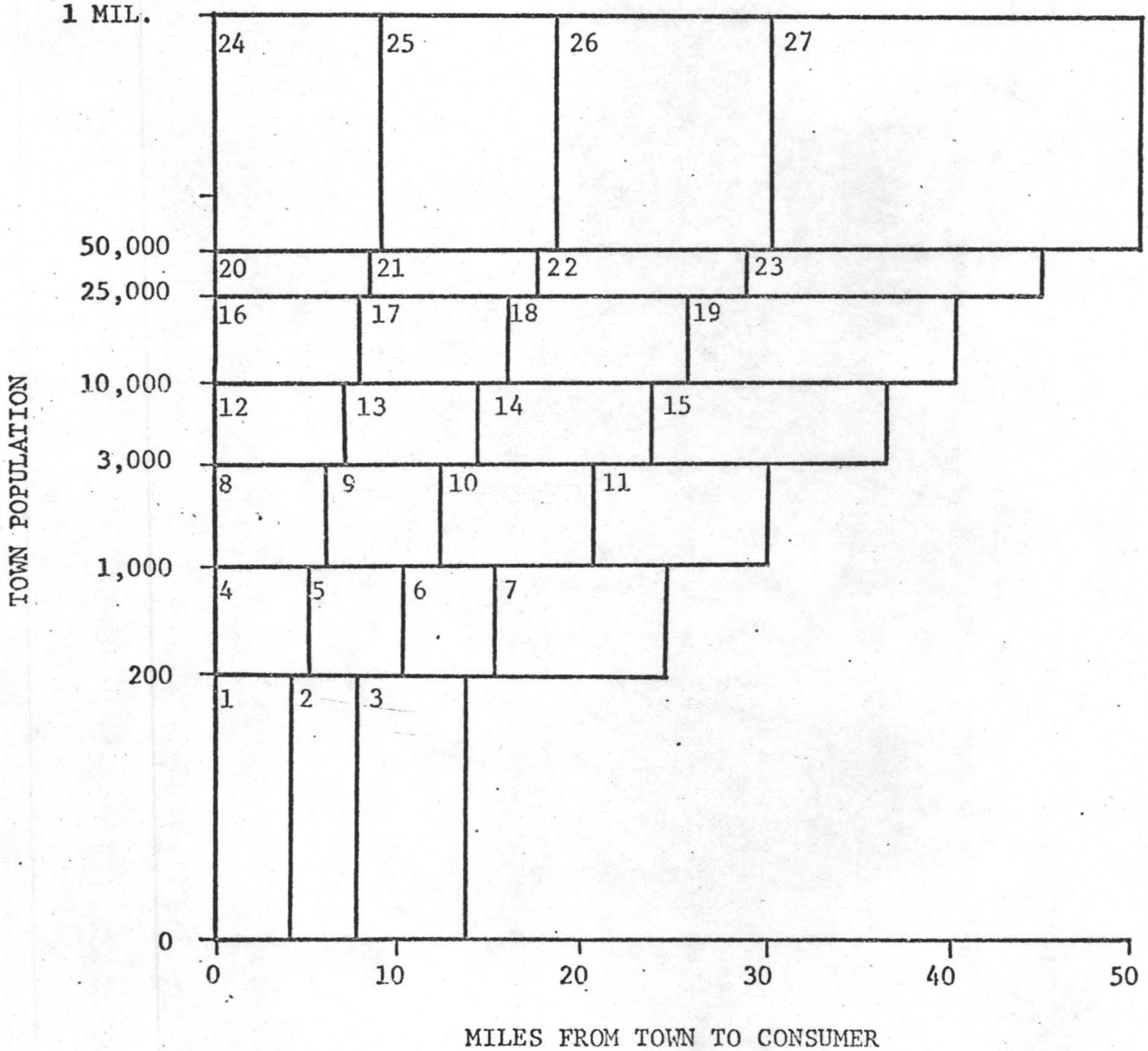
Beyond the information provided by psychological research into the jnd, few other guidelines are available regarding an appropriate scale of analysis. It might be argued that if a hierarchical system of towns existed in Iowa this would provide information regarding appropriate limits for defining "similar" spatial alternatives by town population. Two comments are appropriate. Firstly, at a state-wide level there is little evidence of a hierarchy of towns in Iowa (Goodchild, 1969; Rushton, 1966, p. 29). Secondly, and more basically, even if there were a hierarchical structure, its existence is highly unlikely to have any bearing on the shape of the consumer's preferences. Indeed it is one of this study's basic arguments that preference functions are independent of any particular distribution of alternatives, including a hierarchical distribution. Thus it is erroneous to suggest the distribution should have any bearing on the choice of town population categories used to discover these preferences.

Three criteria, therefore, are used in defining similar spatial alternatives. Firstly, there is the psychological evidence that the larger two things are, the greater their difference must be for people to be able to distinguish between them. Secondly, if two alternatives both possess a similar large amount of one significant attribute, the greater must be their difference on another significant attribute for people to distinguish between them. Thus if two nearby towns are very large, a difference of two or three miles in the consumer's distance to each is unlikely to have the significance it would have with respect to 2 very small towns. Thirdly, there should be sufficiently few spatial alternatives as defined by "town population/distance to consumer" categories, in order that many of the pairs of alternatives may be compared often enough to place reliance on their preference statistic.[5] Using these three criteria, twenty-seven types of spatial

[5]The nature of this statistic will be described in more detail later in the chapter.

FIGURE 3.2

The Definition of Location Types



alternatives are defined (see Figure 3.2), which for simplicity will be called "location types".[6] The criterion for defining the upper distance limit for each town population category is that no town is patronised which is further from the consumer than that upper limit. In other words, no spatial alternative to the right of the total "envelope" of location types is ever chosen.

For the purpose of this analysis, therefore, the simplifying assumption is made that all households choose from some subset of 27 alternative location types. Initially at least, the exact location of any alternative within a location type is irrelevant. Later, however, this information is used.

The Limitations of the Operational Definition of Alternatives

Given the particular definition of location types chosen, it is worthwhile to consider the possible limitations of such a definition and means of testing its usefulness. With respect to the latter goal, Harvey (1966) suggests that, lacking a priori evidence of the appropriateness of a scale of analysis, as in this study, the test of its appropriateness is whether orderliness in a behavioral process is revealed at that particular scale.[7] However, it is very probable that different degrees of order could be discerned in many behavioral processes over a considerable range of scales. Thus proof of the existence of order at one scale would not necessarily indicate the most appropriate

---

[6] The phrase was first used by Rushton (1969a) to define spatial alternatives in terms of the same two attributes as are used here.

[7] Clearly any lack of apparent order might be attributable not only to the inappropriateness of the scale, but also the irrelevance of an independent variable hypothesized to account for behavioral variation.

choice of scale for discerning maximum order. The general import of his remarks is, however, very pertinent to the present analysis, for the ultimate test of the appropriateness of both the independent variables and the scale of analysis used in this study will be the extent to which behavioral order can be discerned.

Within this general frame of reference, however, the scale of analysis does impose certain limitations on the conclusions which can be drawn about the nature of spatial preference. By the very definition of location types, no information is directly provided about consumers' preference between spatial alternatives within the same location type. Preference at a scale any less than that chosen can only be indirectly inferred by assigning the preference "score" of each location type to its centre of gravity,[8] and thereafter interpolating indifference curves between these points, much in the manner of interpolating contours between spot heights. The major weakness of the method in reconstructing both topographic and preference surfaces from a limited number of points is that local irregularities in "relief" are smoothed over. This smoothing would not be likely to affect a test for differences in preference surfaces if differences are relatively pronounced, but if differences are very localised then clearly the scale is too aggregative to pick out these local surface irregularities. Since we have no way of knowing at what scale preference differences are likely to be significant, the analysis can only claim to be a test for differing

---

[8]The centre of gravity is defined as the mean of the co-ordinate locations of all available spatial alternatives within the location type, thus allowing for any non-uniform distribution of available alternatives within the type. The proceedure is suggested as a more accurate aid to reconstructing a preference surface than Rushton's assignment of "scores" to location type mid-points (Rushton, 1969b).

preferences at a scale equal to or greater than the scale used.

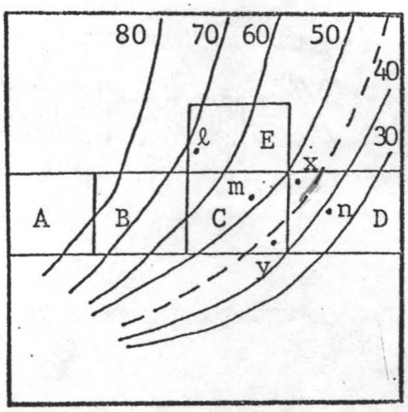
Two other possible limitations due to scale exist. The simpler, alluded to earlier, is that if few comparisons are made between any pair of location types, very large confidence limits must be placed around the preference statistic for the pair. An inordinate number of such cases would make further analysis impossible.[9] Thus one goal of the current definition of location types is to avoid such a situation.

The second, less obvious limitation may result from a combination of the size and <u>shape</u> of the location type. Thus, for example, in Figure 3.3 a hypothetical deterministic preference surface is identified, together with a row of four location types. Using location types C and D as an example, it is clear that despite a deterministic preference surface, C will not be invariably preferred to D. For in the bend between 40 and 50 in C and D, it is possible for a spatial alternative in D to be preferred to one in C, e.g., x is preferred to y. For the most part, however, alternatives in C will be preferred to any in D. The net result is that some proportion of alternatives in C, but not necessarily all, will be preferred to those in D, and therefore a probabilistic statement about preference will be derived. Thus, whilst the shape of the preference surface derived may be a good approximation to reality, preference will be erroneously ascribed probabilistic and not deterministic properties. The example

---

[9] Inevitably, however, certain types of spatial alternatives will be compared very rarely. For example, very few rural households live within 20 miles of two cities of more than 50,000 and as a result, out of 530 households few, if any, comparisons between location types 24 and 25 are to be expected.

FIGURE 3.3



given is an extreme one, insofar as deterministic preference seems unlikely in the light of behavioral research. Nevertheless, the example does highlight the possibility of deriving probability values that suggest greater preference variance than exists,as a result of the size and shape of location types relative to the shape of the preference structure. There is, however, a partial method of testing whether this factor contributes significantly to preference variance. For example, let us assume that the preference surface in Figure 3.3 is a probabilistic one such that for each pair of spatial alternatives l, m and n, l > m, m > n, and l > n on more than half of comparisons made. In other words, it is possible to rank l, m, and n in one dimension, and such a preference ordering is described as being transitive. By contrast, an intransitive preference structure would be where l > m, m > n, but n > l. Now, under the assumption made, this is not the case for l, m, and n. However, even with transitive preference, intransitivity between location types can result from the shape of location types. Thus, even though the true probability of preferring the average alternative in E over the average alternative in C were .6, if slightly more than a tenth of all comparisons between E and C were between alternatives in C lying higher on the preference surface than their paired alternatives in E, then fewer than half the alternatives in E would be preferred to those in C. Likewise, although the average C alternative may be preferred to the average D, D may be preferred to C, given the same conditions described for E and C. However, it is impossible to find any combinations of alternatives in E and D such that the one in D lies higher on the preference surface than one in E. As a result, an intransitive preference structure of the form D > C > E, E > D is possible. Clearly, the closer to 0.5 is the true probability of the average alternative in one location type being preferred to the average in another, the smaller the proportion of cases of the kind described necessary to invert the preference relationship. Thus the total number of intransitivities for all triplets of the 27 location types will provide some indication of the significance of this source of error.[10]

Some Assumptions of the Method

So far, the general method of analysis, the choice of attributes to define alternatives and the operational definition of alternatives have been described. To a considerable extent the efficacy of these choices can be tested in terms of the degree of behavioral order revealed, assuming order exists. However, there are other assumptions made in the analysis which are not empirically testable. At best a rationalization of each can be provided, but there is no ultimate test of their validity.

A fundamental assumption is that if x is chosen more often than y, when both are available, then x is regarded as being preferred to y. However, since x's and y's are defined as location types, it can often happen that more than one alternative in location type x and/or more than one in y are available to the same household. The problem then, is to decide how many paired comparisons have been made. For example, if an x location type is chosen, and

---

[10] Whilst a paucity of intransitivities would suggest that this source of error was not large, the presence of many intransitivities could be attributable to more than this factor. Thus it is only really possible to test the hypothesis that this source of error is relatively slight, but not the alternative hypothesis that it is a significant source of error, since any large number of intransitivities might be consequent upon another source of error, discussed in Chapter 4.

two y's and three z's are not patronised, how many x, y and x, z paired comparisons have been made? Clearly it matters to the calculation of the aggregate proportionate preference statistic for x and z whether in this case x is treated as having been preferred to z three times or once. The assumption here is that only one x,y and one x,z paired comparison has been made, no matter how many alternatives in each rejected location type are available. The behavioral evidence supporting this assumption is provided by Becker, De Groot and Marschak (1963). They describe the problem thus:

> If a person who is offered a choice between a cup of tea and a cup of coffee.....chooses tea more often, we can say he prefers tea to coffee. Suppose this person is approached with a tray on which he sees one cup of tea and two cups of coffee. Is it conceivable that.....he will choose a cup of coffee more frequently than a cup of tea? We feel not. Yet a reasonable model of probabilistically defined preference implies that he might do just that (roughly because if his choices were entirely random, the probability of choosing coffee in the three-cup offer would be twice that of choosing tea).

In an experiment designed to put this paradoxical conclusion to the test, they found that their feeling was supported and that despite the greater number of cups of coffee, coffee was not chosen more frequently than tea.

In this context of choosing between two alternatives, the question also arises whether it is useful and advisable to use as paired comparison information data on the absolute number of dollars a household spends in each town or simply to consider the rank of each town in terms of dollar expenditure. For example, if a household spends $100 in town A, $50 in town B, $25 in town C, and nothing in town D, it is possible to argue that A is preferred to B by the same proportion as dollars are spent in each, i.e., by 100/(100 + 50) or .66; whereas A is preferred to C by 100/(100 + 25) or .8, and A is preferred to D by 100/(100 + 0) or 1., and so on. By contrast, if the rank of each town by expenditure is used, then it is possible only to say, A > B, A > C, A > D, and so on. For two reasons, the latter approach is adopted. Firstly, the absolute dollar expenditure in each town patronized for a given commodity over the span of a year, when recalled the following spring, is likely to be subject to an undetermined amount of error. By comparison, it is probable that recollection of the town in which most was spent on groceries, and of the order of towns in which succeedingly less was spent, is subject to less error. Secondly, most of the paired comparisons between alternatives are between one that was patronised and one that was not. In terms of absolute dollars spent in the one, it is immaterial in this case how much was spent, and therefore the contribution of this piece of paired comparison information in calculating the aggregate proportion of times one location was preferred to the other, is unaffected by whether absolute dollar expenditure or order of expenditure is considered. Considering, therefore, that absolute dollar expenditure information does little to alter the computed proportion of times one location type is preferred to another,[11] and that in any case this kind of data are subject to error,

[11] This was established by analysis. At one decimal point accuracy, about 7% of the paired comparison proportions were different using the two approaches, and even then almost never by more than .1.

there seems no justification for using interval expenditure data when ordinal data produce the same results with less possibility of error.

Finally, it is worth considering the purchasing goal to which the spatial preference between towns is relevant. In theory the purchasing goal is groceries. However, it would be naïve to suggest this is an analysis of urban spatial preference associated with the exclusive purchase of groceries. Inevitably many of the trips made for groceries include purchases of higher order goods which presumably affect the household's choice of town to patronize. In other words, the town chosen is likely to be one satisfying a combination of purchase goals. The significance of the grocery goal in the choice is probably a direct function of the frequency with which groceries are purchased in each of the towns patronized. Thus the more frequently groceries are purchased in a town, the lower is the probability of a higher order purchase being made, since higher order goods are sought less frequently, and hence the more likely is the grocery goal to be the major factor in town choice. But if groceries were purchased as rarely as once a month in a town, it is quite probable that higher order goods would also be sought, which would result in a town being chosen which satisfied higher order goals as well as the grocery goal. And insofar as towns with higher order goods and services than groceries usually provide a similar, if not larger range of grocery outlets than those providing no higher order good than groceries, it is very probable that the grocery goal is as well satisfied in the higher order centre as in centres without the higher order functions. As a result when both groceries and higher order goods are sought, it seems reasonable to assume the major criterion in choosing between towns would be their range and competitiveness in the higher order function, and not their grocery outlets.

No information is available in the data used here on the distribution of purchasing frequency for groceries and other commodities, but it would seem likely that a modal value for grocery purchases for rural households might be once a week for major grocery shopping, whilst fewer but significant numbers of households might make trips twice a week or more at one extreme, and twice a month or less at the other. Irrespective of the precise nature of multi-purpose shopping trips, it is worthwhile emphasizing that the choice rule to be described is not one associated with the exclusive purchase of groceries, or with the purchase of groceries as the highest order good sought but rather it is one in which groceries are a major purchase item. A concern for choice rules associated with the exclusive purchase of groceries or with groceries when purchased as the highest order good sought, whilst of behavioral interest, would be unlikely to be able to explain that significant component of urban choices associated with grocery purchases where higher order goods are also sought. Since central place systems respond structurally to the temporal and spatial interdependencies inherent in people's purchasing varying orders of goods and services, it would be unrealistic not to derive a choice rule explaining urban choices associated with all major grocery purchases.

This chapter has had four purposes:

(a) to describe a method of determining preference from spatial choice, and to demonstrate the logical consistency of the method in contrast to the inconsistency of methods described in Chapter 2;

(b) to describe the data used in the analysis;

(c) to explain the selection of attributes of towns hypothesized to be related to urban preference and to explain the subsequent operational definition of spatial alternatives, i.e., location types; and

(d) to describe and explain the main assumptions underlying the analysis.

The following chapters describe the analyses designed to tackle the problems outlined in Chapter 1.

Chapter 4 THE SHAPE AND PROPERTIES OF THE SAMPLE'S SPATIAL PREFERENCES

In this chapter, the total sample's preference structure is described using the method of paired comparison and inferences are drawn from the revealed pattern of preference. The major part of the chapter, however, is devoted to a description of tests for ordinal and cardinal properties in the aggregate preference structure, and to a discussion of the implications of the test results with respect to:

a) the appropriateness of the attributes chosen to define spatial alternatives;
b) the nature of the sample's spatial choice rules; and
c) the problem of deriving central place systems consistent with ordinal and cardinal spatial choice rules.

The Derivation of the Sample's Aggregate Preference Matrix

In Chapter 3, hypothetical examples are provided of how paired comparison information is obtained from household choices amongst alternatives, and how households' paired comparisons of the same pair are aggregated to provide a statement of the proportion of times one alternative is preferred to another. As indicated in Chapter 3, the data set used provides information on households' spatial choices. The data consist of the Cartesian coordinate locations of all households in the sample, the population and coordinate location of all towns in Iowa with a population of over 50, and the towns patronised by each household for groceries. These are sufficient to calculate the location types chosen and rejected by each household, since the distance from each household to every town can be computed from their coordinate locations.[1,2] A sample of results is provided in Table 4.1. Clearly, for the 27 location types, certain spatial opportunities are more frequently available than others. Thus, for example, large nearby towns, represented by location types 16, 20 and 24 are rarely available, whereas small, middle-distance towns, represented by location types 3, 6 and 7 are numerous. Inevitably, therefore, information from the random sample is uneven.

As indicated in Chapter 3, the choice and rejection of alternatives enables paired comparison information to be derived. For example, the implicit paired comparisons made by the first household in Table 4.1, which patronised location types 14 and 8 and did not patronise location types 2, 3, 6, 7 and 11, are as follows: $14 > 2$, $14 > 3$, $14 > 6$, $14 > 7$, $14 > 8$, $14 > 11$; and $8 > 2$, $8 > 3$, $8 > 6$, $8 > 7$, $8 > 11$. Location type 14 is preferred to 8, since the household allocates more of its grocery dollars to the former, although both are patronised.

This paired comparison information is more simply written in matrix

---

[1]The distance between points is calculated using a city block metric, i.e.,

$$d_{ij} = \sum_{m=1}^{2} |a_{im} - a_{jm}|,$$

where d_{ij} = the distance between points i and j,
a_{im} = the coordinate value of i on the mth dimension of the space, and
m = 1,2 indicates a two-dimensional space.

The Cartesian coordinate system has its origin at a point located approximately 25 miles south and west of the southwest corner of the state of Iowa, and is aligned along a north-south axis. The fact that the road network of Iowa has a strong rectangular grid pattern with a north-south alignment makes the city block metric appropriate, and tests by Rushton (personal communication) indicate a very high correlation between actual and computed road distances.

[2]The programme (LOCTYPE) calculating the location types chosen and rejected by each household is described and listed in Appendix B.

Table 4.1 Number of Towns in Each Location Type Available to Each Household and the Location Types of Towns Patronised, Ranked by Dollar Expenditure

IHD	LOCATION TYPES AVAILABLE 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	LOCATION TYPES PATRONIZED IN ORDER OF $ EXPENDITURE
2	0	1	1	0	0	2	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14 8
3	1	0	5	0	2	1	5	0	1	1	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	10 8
4	0	0	2	0	0	1	3	0	1	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13 5 3
6	0	1	3	0	1	1	5	0	0	2	4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	12
9	0	1	2	0	2	1	4	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	14 0
10	0	1	2	0	1	3	6	0	0	0	3	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	13
11	0	1	1	0	0	3	7	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	14 13 5
12	0	1	3	0	1	3	4	0	0	0	4	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	14
13	0	2	1	0	2	1	7	0	0	0	1	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	6 14 6
14	0	1	1	0	1	2	6	0	0	0	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	13 14
15	0	0	0	0	0	2	5	0	0	0	1	0	0	1	4	0	0	0	0	0	0	0	0	0	0	0	0	5 2 5
16	0	0	0	0	0	0	4	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	5 13 5 2 6
18	0	0	2	0	1	0	7	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	13 0
19	0	0	0	0	0	5	2	0	0	0	1	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	5 2
20	0	1	1	0	0	3	5	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	15 5 15
21	0	1	4	0	1	2	2	0	1	3	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	13 5 3 22
24	0	1	5	0	1	2	3	0	1	0	3	0	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	13 5
25	1	0	3	0	2	0	7	0	0	2	1	0	1	0	2	0	0	0	0	0	0	1	0	0	0	0	0	9
26	0	1	3	0	0	1	4	0	0	1	2	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0	0	8
27	0	3	4	0	0	2	3	0	0	2	2	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	9
28	0	0	1	0	0	1	4	0	0	2	1	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0	0	9
29	0	1	0	0	1	2	4	0	1	2	3	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	22 5
30	0	0	0	0	1	6	6	0	0	1	3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	8
32	0	0	0	0	1	1	7	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9 5
33	0	1	0	0	0	3	7	0	0	3	3	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	4 13
34	0	1	0	0	1	4	10	0	0	1	2	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	22 9 5
35	0	0	3	0	1	1	9	0	0	2	4	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	13 5
36	0	2	4	0	0	4	9	0	1	2	2	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	21
37	0	2	2	0	2	3	7	0	0	2	2	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	21
38	0	0	2	0	0	0	7	0	0	2	3	0	1	1	2	0	0	0	0	0	0	1	0	0	0	0	0	7 4
39	0	0	1	0	3	1	8	0	0	2	2	0	0	1	3	0	0	0	0	0	0	1	0	0	0	0	0	10 5
40	0	0	0	0	2	3	6	0	0	2	4	0	0	0	3	0	0	0	0	0	0	0	1	0	0	0	0	9 14
41	0	1	4	0	2	2	7	0	0	0	5	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	1	13 11 5 10
42	0	1	3	0	1	1	10	0	0	0	1	0	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	6 0
43	0	2	2	0	0	3	8	0	1	3	3	0	0	1	5	0	0	0	0	0	0	0	0	0	0	0	1	5
44	0	1	2	0	1	3	7	0	0	3	3	0	0	1	5	0	0	0	0	0	0	0	0	0	0	0	1	9 6
45	0	2	3	0	1	2	6	0	1	3	3	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	1	5 14
46	1	1	2	0	0	1	9	0	1	0	3	0	1	0	2	0	0	0	0	0	0	0	1	0	0	0	1	13 5 0
47	0	0	1	0	2	3	8	0	0	1	4	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	1	9 13
48	0	0	5	0	1	3	9	0	1	3	2	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	4 9
49	0	1	1	0	1	3	8	0	1	3	3	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	8 9
50	1	1	1	0	1	2	11	0	2	2	3	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	9
52	0	1	4	0	2	4	8	0	1	2	3	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	9 0 15
55	0	0	2	0	1	3	9	0	1	2	3	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	9 9
56	0	1	1	0	1	2	9	1	1	2	4	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	9
57	0	1	3	0	1	3	5	0	0	3	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	1	2
58	0	0	2	0	1	2	6	0	0	4	4	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	9 15
59	0	0	5	0	0	0	5	0	1	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4 0
60	0	1	5	0	1	1	7	0	0	2	3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	9
63	0	0	7	1	0	2	5	0	1	2	6	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1 9 0
64	0	0	7	1	0	2	5	0	1	2	6	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1 9
65	0	0	5	0	0	3	5	0	1	3	3	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	8 4
66	0	2	1	0	0	1	5	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 5 9
67	1	1	2	0	0	4	5	0	0	1	3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	10 1 10 27
68	0	2	3	0	2	1	6	0	1	3	3	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	9

Table 4.2

Number of Times Row Location Type Chosen in Preference to Column Location Type by One Household

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	27
1															
2															
3															
4															
5															
6															
7															
8		1	1			1	1								
9															
10															
11								0							
12															
13															
14		1	1			1	1	1			1				
27															

form for the household (see Table 4.2).[3] Only the lower diagonal section minus the diagonal need be considered since intra-location type comparisons are not considered. Given such a matrix for each household, the sum of matrices indicates the total number of times each row location type is chosen in preference to the column location type. Dividing the sum in each cell by the total number of entries made in that cell (i.e., by the number of observable comparisons made between that pair), provides the proportion of times each location type is preferred to each other by the sample (see Table 4.3). The value -9.9 in this and all succeeding matrices indicates that no information is available on preference between that pair. This can arise either if no household has both alternatives to choose between, or if households do have both to choose between but always choose some other alternative in preference to both. As an indication of the reliability of these sample variates, the sample size upon which each proportion is based is shown in Table 4.4.

Inferences from the Preference Matrix

Given the proportions in Table 4.3, it is useful to note what inferences may or may not be drawn from this statement of aggregate

---

[3] In the rare event of a household spending equal sums in towns belonging to two location types, the appropriate paired comparison score is .5.

Table 4.3

Proportionate Preference of Total Sample for Towns in Row Location Types over Town in Column Location Types

-9.9 SIGNIFIES NO DATA

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	.5																										
2	0.0	.5																									
3	0.0	.1	.5																								
4	.4	.9	1.0	.5																							
5	.4	.8	1.0	.2	.5																						
6	.1	.3	.9	.0	.1	.5																					
7	0.0	0.0	.4	0.0	.0	.0	.5																				
8	1.0	1.0	1.0	.5	.9	1.0	1.0	.5																			
9	.7	.9	1.0	.3	.7	1.0	1.0	.1	.5																		
10	.4	.5	1.0	.0	.2	.8	.9	0.0	.1	.5																	
11	0.0	.1	.5	0.0	.0	.1	.5	0.0	.0	.0	.5																
12	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	.5															
13	.9	.9	1.0	.7	.9	1.0	1.0	0.0	.8	1.0	1.0	0.0	.5														
14	.5	.7	1.0	.2	.5	.9	1.0	.1	.3	.9	1.0	0.0	.0	.5													
15	.1	.2	.9	0.0	.1	.5	.9	0.0	.0	.2	.9	0.0	.0	.1	.5												
16	-9.9	1.0	1.0	1.0	1.0	1.0	1.0	-9.9	1.0	1.0	1.0	-9.9	-9.9	1.0	1.0	.5											
17	1.0	1.0	1.0	.6	1.0	1.0	1.0	.4	.8	1.0	1.0	-9.9	0.0	1.0	1.0	-9.9	.5										
18	0.0	.6	1.0	.4	.4	.8	1.0	0.0	.3	.9	1.0	0.0	.1	.4	1.0	0.0	0.0	.5									
19	0.0	0.0	-9.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-9.9	0.0	0.0	0.0	.5								
20	-9.9	1.0	1.0	1.0	1.0	1.0	1.0	-9.9	1.0	1.0	1.0	-9.9	-9.9	1.0	1.0	-9.9	-9.9	-9.9	1.0	.5							
21	1.0	1.0	1.0	.6	.9	1.0	1.0	.5	.7	1.0	1.0	-9.9	0.0	.9	1.0	0.0	0.0	.8	1.0	-9.9	.5						
22	.7	.9	1.0	.4	.6	.9	.9	.2	.4	.7	1.0	0.0	.2	.6	.9	0.0	.2	1.0	1.0	-9.9	-9.9	.5					
23	0.0	.1	.7	0.0	.1	.3	1.0	0.0	.0	.4	1.0	0.0	0.0	0.0	.4	0.0	0.0	0.0	1.0	-9.9	-9.9	-9.9	.5				
24	1.0	1.0	1.0	1.0	.8	1.0	1.0	-9.9	1.0	1.0	1.0	0.0	.6	1.0	1.0	-9.9	1.0	-9.9	1.0	-9.9	-9.9	1.0	1.0	.5			
25	.7	1.0	1.0	.3	.9	1.0	1.0	.3	.5	.8	1.0	-9.9	.6	1.0	1.0	-9.9	.3	1.0	1.0	-9.9	-9.9	1.0	1.0	-9.9	.5		
26	.5	.7	1.0	.3	.4	1.0	1.0	0.0	.4	.9	1.0	0.0	.1	.2	1.0	-9.9	0.0	.5	1.0	-9.9	.5	-9.9	1.0	-9.9	-9.9	.5	
27	.3	.2	1.0	.2	.1	.9	.9	0.0	.1	.3	.7	0.0	0.0	.1	.8	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	-9.9	-9.9	0.0	.5

Table 4.4

Frequency of Paired Comparisons of Location Types

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	1																										
2	4	14																									
3	16	22	3																								
4	5	22	47	3																							
5	18	77	102	50	83																						
6	17	38	21	52	119	18																					
7	16	27	5	57	116	21	2																				
8	3	26	41	4	36	44	47	1																			
9	20	59	95	36	112	100	113	18	41																		
10	17	31	30	50	99	38	35	33	92	20																	
11	14	25	4	51	103	19	4	41	106	32	2																
12	2	15	25	4	23	30	30	2	3	23	26	0															
13	18	54	91	35	96	103	108	4	40	68	96	3	11														
14	11	54	62	34	104	73	70	30	66	48	65	9	41	25													
15	17	31	14	50	109	28	16	38	96	36	15	23	97	60	13												
16	0	2	8	1	6	9	9	0	3	7	8	0	0	4	5	0											
17	1	11	22	11	21	24	24	5	11	19	22	0	2	9	19	0	0										
18	2	11	9	11	22	13	11	9	19	8	9	3	10	11	9	3	2	2									
19	3	3	0	15	20	3	1	8	22	10	1	6	22	14	0	3	8	3	0								
20	0	3	4	1	3	3	4	0	2	2	4	0	0	3	4	0	0	0	3	0							
21	1	7	12	8	14	14	15	2	11	10	12	0	3	10	15	1	1	4	7	0	0						
22	3	10	11	5	20	14	14	11	16	14	8	5	10	13	10	2	5	4	2	0	0	0					
23	2	8	3	11	33	6	2	12	22	5	2	3	16	17	5	3	12	9	1	0	0	0	0				
24	1	6	5	2	5	6	6	0	4	3	6	2	5	3	6	0	3	0	1	0	0	1	1	0			
25	3	6	10	4	10	9	10	3	6	6	9	0	5	2	9	0	3	4	2	0	0	1	2	0	0		
26	4	10	9	7	21	8	10	5	14	9	9	2	11	9	9	0	3	6	4	0	2	0	3	0	0	0	
27	3	10	6	12	30	7	7	11	28	13	7	9	28	14	5	5	4	4	2	1	7	1	1	0	0	1	0

preference. Firstly, the fact that for some pairs there is considerable disagreement or "confusion" as to which is preferable, i.e., where the proportion is close to .5, does not necessarily indicate that individuals discriminate poorly between the two. It is possible, though not probable, that each individual has a different deterministic preference rule such that he would invariably choose the same one of any two location types, and that the apparently weak discrimination between certain alternatives is simply the result of aggregating dissimilar deterministic rules. Only if the preference matrix were composed entirely of 1's and 0's would it be possible to say all households discriminate perfectly between the same alternatives and therefore that all share the same deterministic preference rule, at least at this scale. With proportions other than 1 and 0, there is no way of telling from the analysis whether the different individual spatial preference rules are deterministic or probabilistic.

For reasons given in Chapter 1, however, it is assumed that preference functions are probabilistic, except perhaps on a broad scale where alternatives may be so different that preference appears deterministic. Assuming preference is probabilistic, the extent to which people do choose differently between the same pair of location types provides critical information about the preference similarity of the location types. The more often one alternative is chosen over another, the further apart they lie in preference space. In other words, the more preferred one alternative is, the less likely is the other to be chosen on any trial. Hence two location types whose preference proportion is close to 1 or 0 are regarded as being very dissimilar in terms of preferability, whilst those with values chose to .5 are treated as close neighbors in preference space.

The proportions also indicate those types of choice situation in which differences in space preferences are likely to occur and those in which such differences are improbable. From the total of 313 pairs of location types for which preference information is available, 174 (55%) of all sample proportions equal 1 or 0, indicating that for these pairs of alternatives, differences in preferences are unlikely. Inevitably, many choice situations are of such a simplistic kind that it would be surprising to find any disagreements in pairwise choices. Such would be comparisons between large nearby towns and small distant ones. However, there are many pairs of location types where one alternative does not score higher in both accessibility and town population, and where one is not invariably chosen over the other. It is in situations such as these that differences in the values placed on the two attributes might result in purposive differences in households' choices from the same pair, producing proportions other than 1 or 0.

If deterministic preference had been assumed, any preference proportion other than 1 or 0 would have been taken to indicate that differing space preferences exist between the households whose choices contribute to that proportion. However, with preference regarded as probabilistic, disagreements in households' choices between the same pair may be attributed to chance and/or to significant and fixed differences in the preference probabilities of different households. Thus a proportion of .52 might indicate that all households have the same preference for one alternative over the other but that it is so slight that imperfect discrimination results in 48% of households choosing the slightly less preferred alternative. Additionally or alternatively it may indicate different probabilistic preference rules which combine to give the proportion .52. For example, the proportion .52 might be decomposable into proportions of .45 and .65 associated with two different types of household. Certainly, the paired comparison matrix in Table 4.3 contains sufficient proportions not equal to 1 or 0 for the hypothesis of significant space preference differences still to be entertained.

<u>Order in the Preference Structure - The Test for an Ordinal Preference Scale</u>

The above discussion concentrates on inferences which can be drawn about preference from the value of specific preference proportions. However, more critical information about lawfulness in preference is available through analysis of the entire preference matrix. The most significant property of the preference matrix is that it indicates whether a simple or complex preference rule is required to explain the proportions in the matrix. Specifically, it indicates whether preference between towns can be described as a simple function of town size and distance. An example of a simple preference rule would be a rank ordering of all location types, in which each type is preferred to all types ranked lower.[4] This property is described as a transitive relation, such that if A > B and B > C, then A > C.

By contrast, a more complex preference structure would contain intransitivities of the type A > B, B > C, yet C > A. As a consequence such a preference relationship could not be represented in one dimension, since C would have to be both lower and higher than A on the scale. In this case two dimensions would be required to represent the two preference orderings embedded in the intransitivity, namely ABC and CAB. Since by definition in this study, individual households cannot have intransitive preferences, any intransitivity would therefore be a consequence of inconsistencies between the preference orderings of households. For example, the above intransitivity could result from the kind of situation described in Figure 4.1, where two major different preference orderings are aggregated. It is to be noted that the intransitivity requires that a majority of the comparisons of A and C run counter to the preferences for A over B, and B over C. Hence inconsistency implies something more than a minority "dissent" in the comparisons of A and C, as would be the case if C were preferred to A only twice in the 5 cases.

Two interpretations of such inconsistencies in households' preference orderings are possible. The more obvious, alluded to already, is that significant numbers of households assign considerably different weightings

---

[4] It should be recalled that one type is considered to be preferred to another if it is chosen more often than the other when both are available.

FIGURE 4.1

The Scaling of an Intransitive Preference Structure

(a) Household Ranking

Household	Ranking
1	ABC
2	ABC
3	CAB
4	CAB
5	BCA

↓

(b) Proportion of times one member of each pair is preferred to the other:

A > B = 4/5

B > C = 3/5

C > A = 3/5

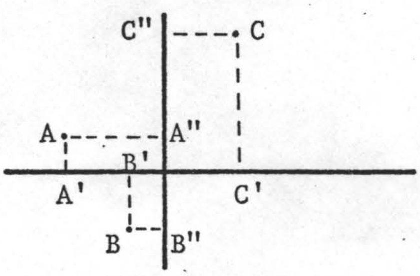
↓

(c) An intransitive preference structure:

A > B > C > A

↓

(d) Two-dimensional scaling of the preference structure, such that the projection of each point onto each axis reveals the two rankings.



to town population and distance, such that the same alternatives are ranked differently. In the economist's terms, there are significantly different trade-offs between town population and distance in households' utility functions.

The second possible interpretation of intransitivities is that one or more major explanatory variables of locational preference have been omitted in the definition of location types. The greater the number of intransitivities, the more significant would be the omissions and the relatively more minor would be the role of the attributes chosen to define location types. Thus to take an extreme example, if the attributes chosen to define location types were quite irrelevant to urban preference, and uncorrelated with relevant ones, a household would be as likely to choose a member of one location type as any other. Thus, each location type would have approximately the same probability (0.5) of being preferred to any other. In such a situation, many intransitivities are likely, since probabilities will be "assigned" to pairs quite randomly and most probabilities by chance will likely be a bit more or less than 0.5. It should be noted, however, that if a preference structure is intransitive, it is not immediately apparent whether one or both of the above explanations is appropriate.

Given that the extent of transitivity provides such information about a preference structure, the total number of intransitivities in the sample's paired comparisons is calculated. The smallest number of alternatives in which an intransitivity can occur is three. A triplet is called a circular triad if its members are intransitively related. The more circular triads there are, the further we depart from a ranking situation. The number of circular triads is expressed as a proportion of the total number of triads of the 27 location types. However not all the implicit paired comparisons derived from households' town choices are eligible for inclusion in a test of transitivity. The condition that an implicit paired comparison must meet is that it be independent of any other paired comparison used in the transitivity test on the triad in question. Since the sample of households is random, the choices and therefore the paired comparisons of different households are regarded as independent of one another. However, certain of the implicit paired comparisons of the same household are not necessarily independent. For example, if a household patronises location type 1 most, and either rejects 2 and 3 or patronises them less, the paired comparisons 1 > 2 and 1 > 3 are not independent events. Clearly the two paired comparisons are dependent on the same event, namely the choice of type 1. An example of how redundant such data are in a test of transitivity is provided in Figure 4.2, where all three alternatives are available to each household and one is patronised by each. No matter what combination of single choices occurs, preference will appear perfectly transitive, simply because of the interdependence of the two paired comparisons derived from each choice, and not necessarily because of any consistency in peoples' preferences.[5]

Where a household ranks two or three of the three alternatives by dollar expenditure, then two out of the three paired comparisons are independent. Thus if the ranking is 1, 2, 3, 1 > 2 and 2 > 3 are independent, since when 1 is ranked first, it is possible for 2 to be preferred to 3 or 3 to 2. However, the paired comparisons 1 > 2 and

---

[5]It should be noted that the method used by Rushton to test for transitivity included the above type of data (Rushton, 1969a).

FIGURE 4.2

Households	Location Types		
	1	2	3
1	1	X	X
2	1	X	X
3	X	1	X
4	1	X	X
5	X	X	1
6	X	1	X
Σ Choices	3	2	1

⟹

	1	2	3
1			
2	2/5		
3	1/4	1/3	

⟹ 1 > 2 > 3

1 = location type chosen.

X = location type available, but rejected.

1 > 3 are not independent since the choice of type 1 determines the form of both paired comparisons.

Therefore, only the following paired comparisons by a household are considered to meet the conditions necessary for the test of transitivity. For the triad being tested for circularity:

(a) if only two of the three alternatives are available to a household and one or both are chosen, then the single paired comparison is acceptable; or

(b) if all three alternatives are available and 2 or 3 are ranked, two of the three implicit paired comparisons are independent, namely the preference for the first over the second, and for the second over the third.

It should be noted that the same choice may satisfy these conditions in one triad but not in another. Thus if a household chooses type 1, rejects 2 and 3 and has no opportunity to patronise a type 4 town, then in the triplet 1, 2, 3, the paired comparisons 1 > 2 and 1 > 3 are not independent and only one can be included in the test. But in the triplet 1, 2, 4, since type 4 is unavailable, the one paired comparison can be included in the test.[6] As a result the entire sample must be analysed for each of the 2196 triplets of 27 location types, and each household's implicit paired comparisons tested for satisfying the above conditions before being included in the transitivity test for that particular triplet. Data are available for all three of the paired comparisons in 1592 out of the 2196 triplets, of which only 38 (2.38%) are intransitive.

Inferences from an Aggregate Ordinal Preference Scale

Several important inferences can be drawn from such a high degree of consistency in households' preference rankings. Firstly, it has been shown that at this scale of analysis, whilst there is not perfect agreement between households' choices in a deterministic sense, there

---

[6] Strictly speaking, the paired comparison 1 > 2 could only be truly independent if no other location types were available. Not only is such a situation almost inconceivable outside of an experimental situation, but the relaxing of the independence requirement to satisfy only the triplet being tested tends to reduce the probability of perfect transitivity rather than increase it. Thus if 2 is preferred to 1 when some other alternative not in the triplet is available and chosen, but 1 is preferred to 2 when that alternative is unavailable, the inclusion of paired comparisons between 1 and 2 when the other alternative is available, will reduce the frequency with which 1 is preferred to 2 and therefore increase the likelihood of intransitivity in triplets containing 1 and 2. Therefore, the lack of perfect independence does not prejudice the result of the test towards perfect transitivity.

is sufficient consistency between households' preferences for it to be possible to represent the general preference structure by one simple ordinal scale[7] (see the order of location types in Table 4.6). This implies orderliness in consumer space preferences, and so reinforces the initial assumption of order in behavior. However, it might be asked how big a difference in households' preferences would result in a higher or lower percentage of intransitivities. An answer is provided by computing the number of intransitivities in artificial samples in which there are two or more different and known preference rules. The same locational co-ordinates for households and towns as in the sample data are used and each household in the artificial data patronises the same number of towns as its real world counterpart. Each household is randomly assigned one of two or more known preference rules and the choices and rejections of towns consistent with that rule are calculated. The information so obtained has the same form as described for the real data in Table 4.1. A test of transitivity is then performed on the entire sample. This test is repeated on sample sets which each have different preference rules. The rules used, and the frequency of each in the sample set, are described in Table 4.5 together with the percentage of intransitivities. Since each rule is some function of one or both attributes defining location types the preference rules are not vastly different. The two most different are the distance minimisation rule and the p/d maximisation rule. Figure 4.3 shows the difference in terms of the indifference surface

---

[7] The method for deriving this scale is described later in the chapter.

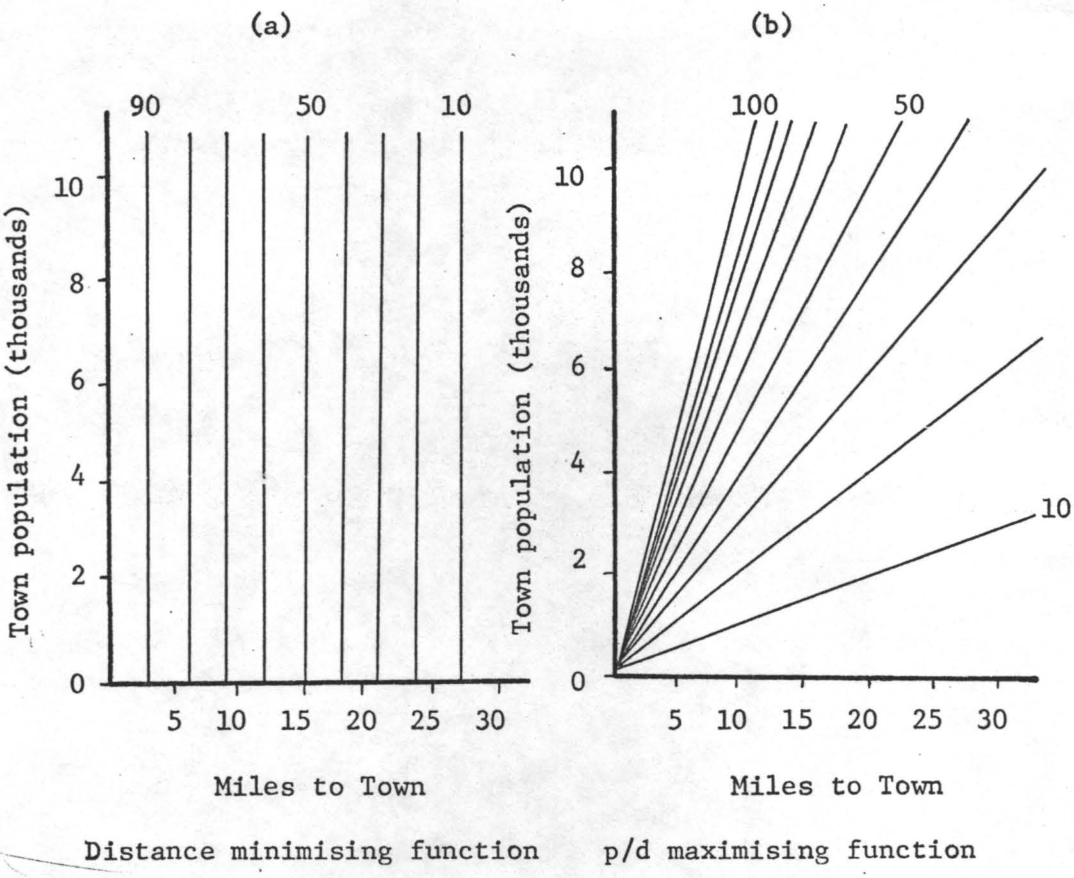
TABLE 4.5

Result of Transitivity Test, Coefficient of Agreement and Result of Ratio-Scale Test for Each Hypothetical Data Set and for the Real Data Set

Preference rules and their proportionate distribution	Deterministic or Probabilistic	% Intransitivity	Coefficient of Agreement	Number of ratio-scale tests	Percentage and number of significant results at: $\alpha = .1$
d minimisation (.5), p/d maximisation (.5)	deterministic	3.40	.67	18	33.0 (6)
p/d^4 max. (.5), p/d^2 max. (.5)	deterministic	1.32	.88	11	0.0 (0)
p/d^3 max. (.5), p/d^2 max. (.5)	deterministic probabilistic	1.09 3.25	.90 .64	11 42	27.0 (3) 21.0 (9)
$p/d^{2.5}$max. (.5), p/d^2 max. (.5)	deterministic	0.74	.90	8	12.0 (1)
p/d^3 max. (.33), $p/d^{2.5}$ max. (.33), p/d^2 max. (.33)	deterministic	0.75	.91	8	12.0 (1)
p/d^3 max. (.1), $p/d^{2.5}$ max. (.1), p/d^2 max. (.8)	deterministic	0.51	.91	8	12.0 (1)
p/d^2 max., normally distributed with d's exponent variance = 1	deterministic probabilistic	0.60 2.32	.83 .56	11 50	18.0 (2) 20.0 (10)
p/d^2 max., normally distributed with d's exponent variance = .5	deterministic	0.50	.88	15	13.0 (2)
Real data set		2.38	.74	12	0.0 (0)

FIGURE 4.3

Indifference Surfaces of Two Hypothetical Preference Functions



characterising each, with ordinal preference values shown.

Whilst there is less than 3% difference between the most and least intransitive preference structures, the direction of difference (see Table 4.5) is as might be expected, given the difference in preference rules and/or the proportions obeying each rule. Thus the differences in percentages of intransitivities, though small, are significant. Comparing the 2.38% intransitivity of the real data set to the percentages for the hypothetical sets, it is clear that all but the first deterministic set in Table 4.5 are more transitive than the real set. The least transitive is not surprisingly the set with the two most different preference rules (see Figure 4.3).

If preference is probabilistic, only one definite statement can be made about the range of preference rules in the real sample. Since it appears that probabilistic rules result in more intransitivities than their deterministic equivalents (see Table 4.5), the 3.4% intransitivity in the most extreme pair of deterministic rules would increase if the rules were probabilistic. Therefore, it is certain that the real sample's distribution of preference rules, even if probabilistic, is less extreme than that set's. Beyond that it is difficult to say what kind of distribution of preference structures the real sample contains. This results both from the limited number of hypothetical probabilistic sets and from the wide range of possible meanings of the word "probabilistic", by comparison with the single interpretation of "deterministic". Thus in the present probability models, each household chooses randomly from the 10 alternatives with the highest preference scores.[8] For example, in the case of p/d^2, the probability of alternative i being chosen by household j is given by the expression:

---

[8] Ten is an arbitrary number used since it seems unlikely that households would look much beyond a 10th most preferable alternative in making a choice. In addition, computer time needed to generate one set of probabilistic data is of the order of 100 minutes and increases as a function of the number of towns ranked by preference score.

$$Pr_{ij} = \frac{(p_i/d_{ij}^2)}{\sum_{i=1}^{10}(p_i/d_{ij}^2)} \tag{4.1}$$

In reality the consumer's probabilistic calculus may be more biased in favor of the larger-appearing alternatives, as defined by their preference scores, than is the case in equation 4.1, i.e., the rule may be more akin to a deterministic one. If so, real preference would be less intransitive than in this model. Or, if larger-appearing alternatives are weighted less, intransitivity would be greater. Thus any conclusions about the sample's range of preference rules, if probabilistic preference is assumed, must be hedged around by "if's". If probabilistic weights are assigned to alternatives as in the hypothetical cases, the sample's distribution of preference rules can be said to be less extreme than p/d^3 max. (.5) and p/d^2 max. (.5), but only marginally wider than p/d^2 max. with d's exponent normally distributed with a variance of 1. If, however, larger-appearing alternatives are more heavily weighted, the real sample's distribution of preference rules may exceed both the above distributions. If the real sample assigns less weight to larger-appearing alternatives than in equation 4.1, then the distribution of rules may be less than the distributions in the two probabilistic sets.

Since none of the hypothetical sets have very different preference rules, the real sample's 2.38% intransitivity suggests that it contains similar rather than identical preference structures. A test of how similar these structures are, is described in Chapter 5.

The second conclusion of major significance is that such a high degree of consistency means that no major explanatory variable has been omitted from the model, and that the two attributes chosen to define location types (and/or their correlates) are the major variables explaining consumer urban choice. As discussed earlier (p. 84), one possible interpretation if more than a few intransitivities were obtained, would be that an attribute with significant bearing on consumer preference had been omitted from the model. In order to corroborate this reasoning the effect of omitting a significant attribute from a preference model was calculated experimentally. A preference model is formulated with 25 location types defined in terms of distance only, namely two-mile categories between 0 and 51 miles. However, households are assigned a preference rule maximising p/d (see Figure 4.3b). The hypothetical data set is generated in the same manner as the others, but although spatial choices are determined by the p/d function to be maximised, they are described only in terms of the distance location types chosen. A test of transitivity reveals 760 intransitivities (32.73%) in 2322 triplets, and since no intransitivity can be attributed to differing preference within this hypothetical sample, the number of intransitivities might normally be higher in a heterogeneous group. The vast difference in intransitivity between this sample with a major preference criterion missing and the real sample (2.38% intransitive) is powerful evidence that no major preference criterion has been omitted from the model. In addition, the relative similarity in intransitivity percentages for the real sample, and for hypothetical samples known to obey preference rules based on the same attributes as define location types, tend to further strengthen this evidence. This would tend to refute the need for a complex model involving more preference criteria, for all but the most detailed central place analyses.

A further conclusion regarding the location types is that the high degree of consistency indicates they are significantly different in terms of their preference distances. David (1963, p. 21) puts it thus:

> if there is no difference between the objects the [judges] cannot reasonably be expected to be consistent, while it is easy for [them] to be consistent if the differences are great.

Clearly, differences are sufficient for a high degree of consistency to be found, but are not so vast that in every paired comparison there is perfect agreement as to which location type is preferrable.

In addition to the inferences drawn above, proof of an ordinal preference scale also has an implication for future data collection. Knowing the general order of preference, future samples should be biassed towards choices between location types close together in the ordering, where present sampling error still casts doubt on the exact ordering. Consequently, samples should avoid trivial choice situations where one location type is so very much higher on the ordering than the others available that the former's ordinal position relative to the others is almost certainly unaffected by sampling error.

One other measure of orderliness in preference sheds light on the extent of <u>agreement</u> in preferences. The coefficient of agreement as formulated by Kendall(1955, pp. 148-149) is based on the proportion of all paired comparisons which are the same. The expression is as follows:

$$\frac{2\Sigma\binom{\gamma}{2}}{\Sigma\binom{m}{2}} - 1 \qquad (4.2)$$

where γ = the number of similar choices in the same pair,

m = the number of times a pair is compared, and

Σ = is the sum over all off-diagonal cells in the triangular paired comparison matrix.

If γ equals m in all cells, there is 100% agreement between households in every paired comparison and the coefficient equals one. If γ equals $\frac{m}{2}$ in every cell, there is maximum disagreement (i.e., 50%) in all paired comparisons and the coefficient equals zero. The coefficient for the total sample is .74 indicating 87% agreement. However, the coefficient varies according to the preference similarity of alternatives, with disagreement more likely between preferences for similar things. Hence the value is scale dependent, in that a grosser location type scale would result in a higher coefficient and a finer scale in a lower coefficient. Thus the figure .74 has limited interpretability unless compared with other coefficients derived at the same scale and from households with the same spatial alternatives. Using the artificial data, such a comparison is possible since the scale is the same and the hypothetical behavior is based on the same household and town locations. The results are shown in Table 4.5. Clearly the sample is considerably more in agreement than any of the two probabilistic sets and less in agreement than all but one of the deterministic sets. Since this result is similar to the findings regarding the real sample's transitivity relative to the hypothetical samples, similar conclusions can be drawn regarding the sample's distribution of preference structures.

Derivation of the Ordinal Preference Scale

For the test of transitivity it is necessary to use only independent paired comparisons. However, given the proof of transitivity, the same restriction is unnecessary in deriving the form of the location type ranking. Essentially a preference ranking of two things, say A and B, is simply a statement of which is more often chosen when both are available, and it is immaterial whether or not such information is obtained from the same choices that indicate whether A is preferred to C. Therefore the ranking of the 27 location types is derived using all implicit paired comparisons contained in households' choices.

The method of obtaining the most appropriate ranking is more straightforward in the case where no data are missing from the preference matrix. Therefore, this method is described as a prelude to explaining the choice between alternative methods if data are missing. Without missing data, a matrix of the type in Table 4.3 is rewritten as a rectangular matrix such that $\hat{p}_{ij} = 1 - \hat{p}_{ji}$ for all i and j. Rows and columns can be rearranged such that the first row has most $\hat{p}$ values greater than .5. All other rows and columns can be similarly rearranged so that location types are ranked in order of preference from top to bottom and left to right in the matrix (see Table 4.6). This provides the ranking most in accord with the preference probabilities. Also the number of intransitivities in the matrix can be easily computed without comparing the probabilities in all triplets. Kendall (1955, p. 148) provides the formula for computing d, the number of intransitivities:

$$d = \frac{1}{12} n(n - 1)(2n - 1) - \frac{1}{2} \sum_{j=1}^{n} a_j^{\,2}, \qquad (4.3)$$

Table 4.6

Proportionate Preference of Total Sample for Towns in Row Location Types over Towns in Column Location Types: Location Types Ranked in Order of Preference from Top to Bottom and Left to Right

	12	16	20	24	8	17	13	21	25	4	9	22	1	26	14	5	18	2	10	27	6	15	23	3	11	7	19	aj 26
12	.5	-9.9	-9.9	1.0	1.0	-9.9	1.0	-9.9	-9.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	25
16	-9.9	.5	-9.9	-9.9	-9.9	-9.9	-9.9	1.0	-9.9	1.0	1.0	1.0	-9.9	-9.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	24
20	-9.9	-9.9	.5	-9.9	-9.9	-9.9	-9.9	-9.9	-9.9	1.0	1.0	-9.9	-9.9	-9.9	1.0	1.0	-9.9	1.0	1.0	1.0	1.0	1.0	-9.9	1.0	1.0	1.0	1.0	23
24	0.0	-9.9	-9.9	.5	-9.9	1.0	.6	-9.9	-9.9	1.0	1.0	1.0	1.0	-9.9	1.0	.8	-9.9	1.0	1.0	-9.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	22
8	0.0	-9.9	-9.9	-9.9	.5	.6	1.0	.5	.7	.5	.9	.8	1.0	1.0	.9	.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	21
17	-9.9	-9.9	-9.9	0.0	.4	.5	0.0	1.0	.7	.6	.8	.8	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	20
13	0.0	-9.9	-9.9	.4	0.0	1.0	.5	1.0	.4	.7	.8	.8	.9	.9	1.0	.9	.9	.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	19
21	-9.9	0.0	-9.9	-9.9	.5	0.0	0.0	.5	-9.9	.6	.7	-9.9	1.0	.5	.9	.9	.8	1.0	1.0	1.0	1.0	1.0	-9.9	1.0	1.0	1.0	1.0	18
25	-9.9	-9.9	-9.9	-9.9	.3	.3	.6	-9.9	.5	.3	.5	1.0	.7	-9.9	1.0	.9	1.0	1.0	.8	-9.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	17
4	0.0	0.0	0.0	0.0	.5	.4	.3	.4	.8	.5	.7	.6	.4	.7	.8	.8	.6	.9	1.0	.8	1.0	1.0	1.0	1.0	1.0	1.0	1.0	16
9	0.0	0.0	0.0	0.0	.1	.2	.2	.3	.5	.3	.5	.6	.7	.6	.7	.7	.7	.9	.9	.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	15
22	0.0	0.0	-9.9	0.0	.2	.2	.2	-9.9	0.0	.4	.4	.5	.7	-9.9	.6	.6	1.0	.9	.7	1.0	.9	.9	-9.9	1.0	1.0	.9	1.0	14
1	0.0	-9.9	-9.9	0.0	0.0	0.0	.1	0.0	.3	.6	.3	.3	.5	.5	.5	.6	1.0	1.0	.6	.7	.9	.9	1.0	1.0	1.0	1.0	1.0	13
26	0.0	-9.9	-9.9	-9.9	0.0	0.0	.1	.5	-9.9	.3	.4	-9.9	.5	.5	.2	.4	.5	.7	.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	12
14	0.0	0.0	0.0	0.0	.1	0.0	.0	.1	0.0	.2	.3	.4	.5	.8	.5	.5	.6	.7	.9	.9	.9	.9	1.0	1.0	1.0	1.0	1.0	11
5	0.0	0.0	0.0	.2	.1	0.0	.1	.1	.1	.2	.3	.4	.4	.6	.5	.5	.6	.8	.8	.9	.9	.9	.9	1.0	1.0	1.0	1.0	10
18	0.0	0.0	-9.9	-9.9	0.0	0.0	.1	.3	0.0	.4	.3	0.0	0.0	.5	.4	.4	.5	.6	.9	1.0	.8	1.0	1.0	1.0	1.0	1.0	1.0	9
2	0.0	0.0	0.0	0.0	0.0	0.0	.1	0.0	0.0	.1	.1	.1	0.0	.3	.3	.2	.4	.5	.5	.8	.7	.8	.9	.9	.9	1.0	1.0	8
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	.2	.0	.1	.3	.4	.1	.1	.2	.1	.5	.5	.7	.8	.8	.6	1.0	1.0	.9	1.0	7
27	0.0	0.0	0.0	-9.9	0.0	0.0	0.0	0.0	-9.9	.2	.1	0.0	.3	0.0	.1	.1	0.0	.2	.3	.5	.9	.8	1.0	1.0	.7	.9	1.0	6
6	0.0	0.0	0.0	0.0	0.0	.0	.0	0.0	0.0	.0	.0	.1	.1	0.0	.1	.1	.2	.3	.2	.1	.5	.5	.7	.9	.9	1.0	1.0	5
15	0.0	0.0	0.0	0.0	0.0	0.0	.0	0.0	0.0	0.0	.0	.1	.1	0.0	.1	.1	0.0	.2	.2	.2	.5	.5	.6	.9	.9	.9	-9.9	4
23	0.0	0.0	-9.9	0.0	0.0	0.0	0.0	-9.9	0.0	0.0	.0	-9.9	0.0	0.0	0.0	.1	0.0	.1	.4	0.0	.3	.4	.5	.7	1.0	1.0	1.0	3
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	.0	0.0	0.0	0.0	.0	.0	0.0	.1	.0	0.0	.1	.1	.3	.5	.5	.6	-9.9	2
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	.0	0.0	0.0	0.0	0.0	.0	0.0	.1	.0	.3	.1	.1	0.0	.5	.5	.5	1.0	1
7	0.0	0.0	0.0	0.0	0.0	0.0	.0	0.0	0.0	0.0	.0	.1	0.0	0.0	.0	.0	0.0	0.0	.1	.1	.0	.1	0.0	.4	.5	.5	1.0	0
19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-9.9	0.0	-9.9	0.0	0.0	.5	

where a_j = the number of $\hat{p}$ values in row j greater than .5. Where the matrix is complete, the number calculated by the formula equals the actual number of intransitivities. However, where data are missing, the number calculated can exceed the actual number, and unlike the case without missing data, the number is not invariant over permutations of rows and columns (Coombs, 1964, p. 358). Thus the most appropriate ranking with missing data is that which minimises the number of intransitivities calculated by Kendall's formula.[9]

Four ranking algorithms were tested on sample preference matrices with missing data. The following are the criteria for ranking location type i:

(1) $\sum_{j=1}^{27} \hat{p}_{ij}$,

(2) $\sum_{j=1}^{27} (\hat{p}_{ij} \times n_{ij})$,

(3) $\sum_{j=1}^{27} (\hat{p}_{ij} > .5)$, and

(4) compare $\hat{p}_{ij}$ and $\hat{p}_{kj}$, j = 1, 27, and rank location type i higher than k if $\hat{p}_{ij} > \hat{p}_{kj}$ more often than $\hat{p}_{kj} > \hat{p}_{ij}$.

Comparison of the results obtained using the four alternative algorithms show the first criterion to produce consistently fewer intransitivities

---

9 Since Kendall's formula can only be used in a complete matrix, blank cells above the diagonal of the matrix are assumed to have values greater than .5 and those below to have values less than .5. Values on the diagonal itself are ignored in the calculation, since they represent paired comparisons of location types with themselves.

than any other method.

To avoid assigning too low a rank to location types with missing data cells, and whose $\sum_{j=1}^{27} \hat{p}_{ij}$ is therefore decreased, the ranking algorithm, in comparing each pair of location types, sums only those $\hat{p}_{ij}$ and $\hat{p}_{kj}$ where information is available for both. The rectangular ordered preference matrix obtained using this ranking procedure upon the initial preference matrix, is shown in Table 4.6. The ranking of location types according to their a_j score (see Table 4.6) is shown in Figure 4.4. The trade-off between the utilities of accessibility and population is clear from the ranking. However, other implications of the ranking are more appropriately discussed after the test for a cardinal preference scale is described.

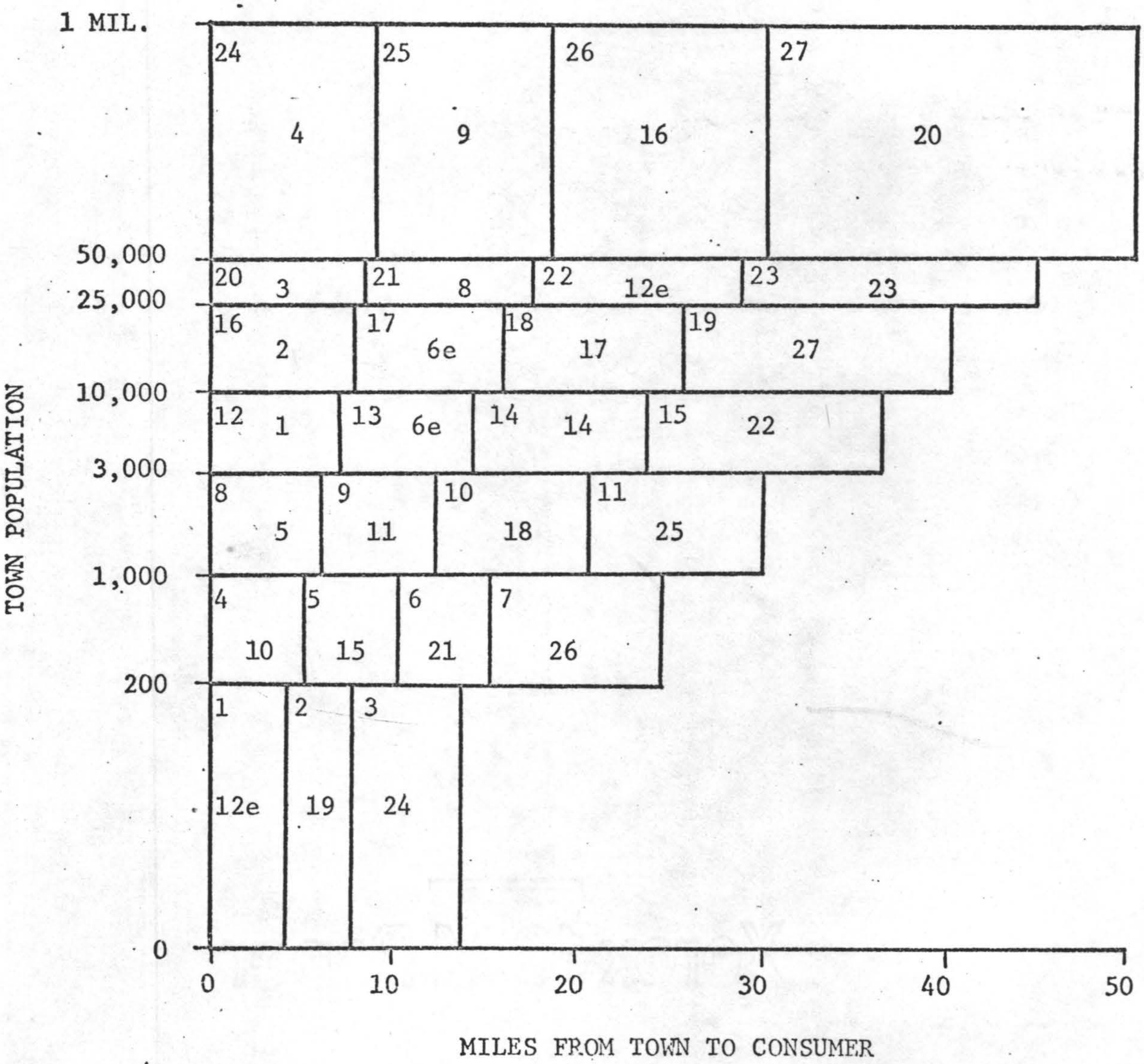
<u>Consumer Preference Functions Appropriate for Testing Central Place Theory</u>

Before testing for cardinal properties in the preference structure, it is worthwhile considering what type of consumer preference postulate would be most useful to theoretical central place analysis and to compare an ordinal preference rule with this ideal.

The basic hypothesis of Central Place Theory is that the number, size and spacing of central places on a demand surface is such that the central place system comes as close as possible to maximising people's spatial preference functions, within the threshold constraints imposed by the suppliers' profit goal. In a market economy, such a system is arrived at by each centre competing freely with every other in an attempt to maximise its share of the market, within the constraint imposed by the form of the consumer's spatial preference function. The ultimate pattern presumably reflects an equilibrium adjustment, or an approximation

Figure 4.4

The Sample's Preference Ranking of Location Types



1 denotes the most preferred location type.

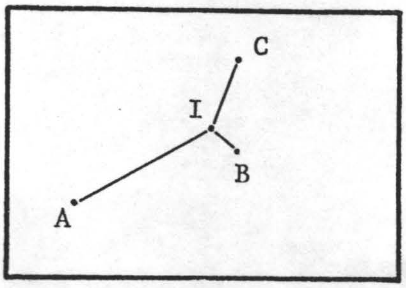
thereof, of spatial supply to demand. To have a testable model of this situation, one prerequisite is that for any arrangement of central places it should be possible to compute for each point on the demand surface the probability of a household or households at that point interacting with each competing central place in the system. Only by this means, is it possible to compute the most probable income of each supplier and hence the viability of each central place, as well as compute the total income of suppliers which is maximised in an optimal system. Thus for example, in the situation illustrated in Figure 4.5, it should be possible to compute what proportion of expenditures by consumers at I can be expected to be allocated to each of the three centres and to any other competitive centres not shown. However, an ordinal preference function, which is necessarily deterministic, can only predict that all expenditures from one place will go to the same place, despite the indications that urban spatial preference is probabilistic rather than deterministic.[10] But as the data indicate (Table 4.1) many households patronise more than one place for a given good, and though similarly located with respect to a set of alternatives, households frequently allocate differing proportions of their expenditure to each centre. But the present ordinal preference model can only predict that all households at the same point allocate all of their expenditures to the same centre.

---

[10] Only for the rare cases where only two places are at all likely to obtain a consumer's patronage, do we have the information to compute his probable allocation to each place, in the form of pairwise preference probabilities.

Figure 4.5

A Hypothetical Spatial Choice Situation



In terms of deducing a realistic central place system, ordinal preference postulates may well produce spurious results. For example, consider the town A in Figure 4.5 in competition with other towns, B and C, for the patronage of consumers in the area. Now if A belongs to some location types which are only marginally preferred to some of those to which B and C belong, and if A also belongs to other location types which are almost invariably preferred by location types to which B and C belong, then it is quite possible that A may not be a viable centre. However, an ordinal preference rule which says nothing about the keenness of competition indicated by preference probabilities would suggest A was just about as often preferred to B and C, as they were to A. Consequently, the ordinal preference rule would give A a much better chance of remaining in the system than the more realistic probabilistic rule. However, the problem is that to predict the probability of one alternative being chosen from more than two alternatives, given only pairwise probabilities, requires proof that a ratio scale of preference exists.[11]

[11] The justification of this assertion is provided in the next section.

Order in the Preference Structure - the Test for a Ratio Preference Scale

Proof of the existence of a ratio scale of preference depends on proof that preference probabilities conform to two assumptions of conditional probability. The first is that the probability of an event, in this case, the choice of one location type from a set, is constant between trials, and the second is that trials, in this case, paired comparisons, are independent. If pairwise preference probabilities meet these conditions, then it is valid to use these probabilities to calculate conditional probabilities of the form $P\{x|T\}$, defining the probability of location type x being chosen from all location types. It is these latter probabilities which can be used to define ratio scale values for location types, representing their amount of "preferredness" or utility.

The second of the two assumptions is readily satisfied by meeting the same requirement for independence between paired comparisons as was specified for the transitivity test (see page 86 for details). The empirical validity of the constant probability assumption, however, is less readily verified but can be checked indirectly by testing whether a consequence of the assumptions is valid. Luce (1959) has shown that a consequence of the assumptions in conditional probability is that one pairwise preference probability can be estimated from two others. Since this calculated value can be compared with an observed one, a test is provided of the empirical validity of using conditional probability and its assumptions to describe preference. It is the result of this test, therefore, which provides information on whether a ratio scale of preference exists. Details of the test are shown below.

The method of calculating one pairwise probability from two others is based on the definition of conditional probability:

$$P\{x|R\} = P\{x \cap R\}|P\{R\} \tag{4.4}$$

where $P\{x|R\}$ is the conditional probability of event x, given event R. In terms of choice, this defines the probability of choosing location type x, given that a type from set R is chosen. However, in a choice situation, x is defined as contained in R, i.e.,x is one of two or more alternatives in set R, and therefore (4.4) can be rewritten as:

$$P\{x|R\} = P\{x\}|P\{R\} \tag{4.5}$$

since $P\{x\} \equiv P\{x \cap R\}$ where $x \subset R$.

Cross-multiplying (4.5) gives:

$$P\{x\} = P\{x|R\}.\ P\{R\} \tag{4.6}$$

If R is a set containing only the pair of alternatives x and y, (4.6) can be rewritten as:

$$P\{x\} = P\{x|x \cup y\}.\ P\{x \cup y\} \tag{4.7}$$

Likewise $P\{y\} = P\{y|x \cup y\}.\ P\{x \cup y\} = [1-P\{x|x \cup y\}]\ .\ P\{x \cup y\}$ (4.8)

since $P\{x|x \cup y\} + P\{y|x \cup y\} = 1$.

$$\therefore\ \frac{P\{x\}}{P\{y\}} = \frac{P\{x|x \cup y\}}{1-P\{x|x \cup y\}} \tag{4.9}$$

Similarly, $$\frac{P\{y\}}{P\{z\}} = \frac{P\{y|y \cup z\}}{1-P\{y|y \cup z\}} \tag{4.10}$$

and $$\frac{P\{x\}}{P\{z\}} = \frac{P\{x|x \cup z\}}{1-P\{x|x \cup z\}} \tag{4.11}$$

(4.11) can be rewritten thus:

$$\frac{P\{x|x \cup z\}}{1-P\{x|x \cup z\}} = \frac{P\{x\}}{P\{y\}} \cdot \frac{P\{y\}}{P\{z\}} \tag{4.12}$$

Substituting the right-hand sides of (4.9) and (4.10) in (4.12) gives:

$$\frac{P\{x|x \cup z\}}{1-P\{x|x \cup z\}} = \frac{P\{x|x \cup y\}}{1-P\{x|x \cup y\}} \cdot \frac{P\{y|y \cup z\}}{1-P\{y|y \cup z\}} \tag{4.13}$$

Defining this ratio as r:

$$P\{x|x \cup z\} = r/(1+r) \tag{4.14}$$

Since r is computed using only the two probabilities $P\{x|x \cup y\}$ and $P\{y|y \cup z\}$, one pairwise conditional probability has been calculated from two others. The test of the validity of this operation on preference probabilities is whether the independently observed sample value of $P\{x|x \cup z\}$ is not significantly different from this theoretical value. If it is not, then the constant probability assumption can be considered valid.

Letting $p_{ik} = P\{i|i \cup k\}$, the theoretical value of p_{ik} can be calculated as a function of p_{ij} and p_{jk}, and compared to the independently observed sample value $\hat{p}_{ik}$.

Specifically, the hypotheses tested are:

$$H_0 : p_{ik1} = p_{ik2}$$

$$H_1 : p_{ik1} \neq p_{ik2}$$

where p_{ik1} is the observed probability of i being chosen over k, when one is chosen. The estimate of this ($\hat{p}_{ik1}$) is based on the number of times type i is chosen (x_{ik1}) in n paired comparisons of i and k, i.e., $\hat{p}_{ik1} = x_{ik1}/n_{ik1}$. P_{ik2} is the theoretical probability of the same event, based on the two probabilities p_{ij} and p_{jk} of which we have estimates $\hat{p}_{ij}$ and $\hat{p}_{jk}$.

A Z test is used to compare the two sample proportions, for all cases where the sampling distribution of the binomial random variables is approximately normal, i.e. whenever a proportion is based on a sample size $\geq$ 30 and np and n(1-p) $\geq$ 5. However, little efficiency in the test is lost by relaxing sample size conditions to those for the χ^2 test. Thus, all cases are tested where the sum of the sample size for the observed proportion and the smaller of the two for the computed proportion is greater than or equal to 20, and np and n(1-p) $\geq$ 5 for all three proportions.

The usual Z statistic used in comparing two sample proportions takes the form:

$$Z = \frac{\hat{p}_{ik1} - \hat{p}_{ik2}}{\sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}}} \qquad (4.15)$$

where p, the pooled estimate of both sample proportions is defined as $(x_{ik1} + x_{ik2})/(n_{ik1} + n_{ik2})$. However, since p_{ik2} is computed from two other proportions, the basic Z statistic requires some modification. In particular x_{ik2} and n_{ik2} are undefined, so that a pooled estimate, p, cannot be defined.[12] Also, in calculating the variance terms in the denominator for $\hat{p}_{ik1}$ and $\hat{p}_{ik2}$, the latter's variance is some function of the variances of $\hat{p}_{ij}$ and $\hat{p}_{jk}$. To allow for this, (4.15) is rewritten as:

$$Z = \frac{\hat{p}_{ik1} - \hat{p}_{ik2}}{\sqrt{\dfrac{\hat{p}_{ik1}(1-\hat{p}_{ik1})}{n_{ik1}} + \mathrm{Var}\,(\hat{p}_{ik2})}} \tag{4.16}$$

where a close approximation to $\mathrm{Var}(\hat{p}_{ik2})$ is given by:

$$\mathrm{Var}\,(\hat{p}_{ik2}) = \frac{\dfrac{\hat{p}_{ij}(1-\hat{p}_{ij})}{n_{ij}} \cdot \dfrac{\hat{p}_{jk}(1-\hat{p}_{jk})}{n_{jk}} + \hat{p}_{ij}^{\,2}\,\dfrac{\hat{p}_{jk}(1-\hat{p}_{jk})}{n_{jk}} + \hat{p}_{jk}^{\,2}\,\dfrac{\hat{p}_{ij}(1-\hat{p}_{ij})}{n_{ij}}}{(\,1 + 2\hat{p}_{ij}\hat{p}_{jk} - \hat{p}_{ij} - \hat{p}_{jk})^2} \tag{4.17}$$

Using the above statistic, the results in the twelve cases satisfying the sampling constraints are shown in Table 4.7. With $\alpha = .1$ and

---

[12] The effect of using two variance terms not based on the pooled estimate, p, is to reduce the value of the denominator in (4.15) when n_{ik1} and n_{ik2} are similar and therefore increase the Z statistic and with it, the likelihood of rejecting the null hypothesis. If n_{ik1} and n_{ik2} are different, the absence of a pooled estimate is as likely to decrease as increase the denominator and therefore cannot be thought of as increasing the likelihood of accepting the null hypothesis.

confidence limits of $-1.64 \leq Z \leq 1.64$ all tests accept the null hypothesis.

The implication of this result is that the method used to compute one probability from two others is valid in the context of preference, as are the assumptions on which the method is based. In particular, the assumption of the constant probability of an event is verified. This means that the probability of choosing one location type over another is invariant under changes in the other alternatives available, which changes are not uncommon in the data analysed. This in itself is suggestive of a ratio scale of preference on which the relative distances between location types are invariant. But the most important point to note is that all probabilities are implied to be constant, including $p\{x|T\}$ for all $x \subset T$. If so, each location type x can be assigned a unique score, v(x), equal to $kp\{x|T\}$, where k is any constant. Since $p\{x|T\}$ has a range of 0 to 1, any multiplication by k leaves the origin unchanged at 0 and simply "stretches" the scale uniformly without altering the relative v scores of location types. This is the definition of a ratio scale.

Returning to the fact that only 12 cases were tested, since the 10 location types involved are spread over a considerable range of the scale, this is considered sufficient to infer that the entire scale has ratio properties. The only parts of the scale not represented by the 10 location types (2, 4, 5, 6, 9, 10, 13, 14, 19, 22) are the upper and lower extremes, where location types are either too infrequently compared to others to meet the sample size requirements

TABLE 4.7

Results of Ratio-Scale Test on Real Sample

z	$\hat{p}_{ik1}$(observed)	$\hat{p}_{ik2}$(estimated)	Location types i	k	j	n_{ik}	n_{ij}	n_{jk}
-.750	.467	.600	10	2	5	15	70	36
-.526	.727	.805	14	2	5	22	67	50
-.687	.758	.876	14	2	10	33	30	29
1.053	.476	.318	9	4	5	21	120	41
.048	.565	.555	13	4	5	23	104	51
.291	.629	.582	13	4	9	35	39	34
-.967	.179	.281	14	4	9	28	70	28
-.001	.200	.201	14	4	13	35	43	32
.829	.471	.331	10	5	6	17	16	49
.990	.832	.659	13	5	9	95	16	123
-.980	.488	.627	14	5	9	86	35	81
-.954	.438	.678	22	5	9	16	10	121

of the test, or are so frequently preferred to or by other alternatives that np and n(1-p) < 5. Thus, the absence of z tests for such location types is the effect of sampling constraints rather than evidence that the preference scale has interval properties over only part of its range.

Inferences from an Aggregate Ratio Preference Scale

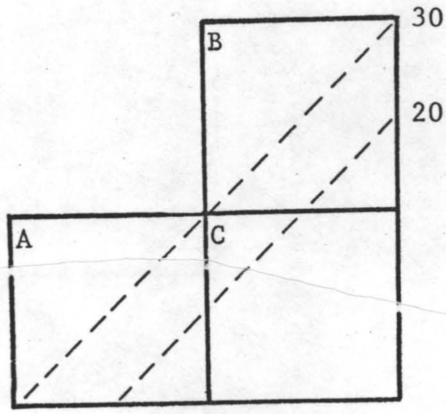
In order to ascertain the significance of the inference made from the z test results, the test was run on the same hypothetical samples as were tested earlier for transitivity. Details of the results are given in Table 4.5.

The very small number of tests in all but the two probabilistic sets makes inference tenuous. However, it appears that in general the narrower the distribution of preference rules in a sample set, the lower is the percentage of statistically significant differences. Thus, the likelihood of an aggregate ratio scale of preference would seem to be a function of the homogeneity of the preference rules contained in a sample. Also, the likelihood of a ratio scale would not appear to be greater if individuals' preference rules are probabilistic rather than deterministic.

The fact that several deterministic preference scales apparently have ratio properties is highly critical and seems most likely to be a function of the scale of the analysis. In Figure 4.6 two indifference curves describe the general shape of a deterministic preference rule as it refers to the three location types A, B and C. Now, whilst an alternative is invariably preferred to all others lower than it on the indifference surface, it is clear that no location type can be invariably preferred to another since each contains alternatives which are both lower and higher on the surface than alternatives in any other location type. In fact, the average proportion of times A is preferred to B can be obtained by comparing the relative proportions of each location type lying above and below the same indifference curve. This assumes a uniform distribution of alternatives in each location type. But, even if this assumption is not met, the same conclusion can be drawn provided the distribution of alternatives in A, for example, is the same when compared to B as when compared to C. Thus, the probability of A being preferred to B (p_{AB}) is .5; p_{AC} = .875; and p_{BC} = .875. Without troubling to use Luce's formulation, it is clear one probability can be computed from the other two. For, if there is indifference between A and B then on a ratio scale A should be preferred to C by the same amount as B is preferred to C. Thus, proof of ratio properties in the <u>aggregate</u> preference scale does not necessarily indicate that <u>individual</u> preference scales have ratio properties.

Nevertheless, proof of an aggregate ratio scale does have utility on other levels. One advantage is that knowledge that a ratio scale

FIGURE 4.6



exists is sufficient to calculate the ratio scores of all location types, and hence all n(n-1)/2 paired comparison proportions, from a minimum of n-1 proportions. In the present analysis this is not done due to the considerable sampling error in many proportions. Also, the large number of proportions available in this study permits a more reliable interval scale to be calculated using a multidimensional scaling algorithm described later in the chapter. However in future studies, by concentrated sampling on only nearby location types on a preference scale, large samples would enable all location types' ratio scores to be calculated using the formula:

$$v(x) = \frac{v(y) \cdot P\{x \mid x \cup y\}}{1 - P\{x \mid x \cup y\}} \qquad (4.18)$$

given that an initial arbitrary value was assigned to one location type from the entire set.

The effort saved in not sampling all n(n-1)/2 paired choice situations, most of which contain redundant information, could be used in more intensive sampling of the few paired choice situations required. The saving could also be used to reduce the range of population/distance combinations covered by each location type, thus reducing the effect of aggregation.

In addition, a ratio scale is more useful in testing central place hypotheses than an ordinal one, if, as the data indicate, households patronise more than one centre. An ordinal scale can only predict a household to patronise one place. But with a ratio scale, the household's probability of patronising any one alternative in a set can be estimated using only the original pairwise probabilities. Luce (1959) provides the equation:

$$P\{x|S\} = 1 \bigg/ \sum_{y \in S} \frac{P\{y|x \cup y\}}{P\{x|x \cup y\}} \qquad (4.19)$$

where S is any subset of location types in the total set T.

Consequently, it is possible to specify what proportion of expenditure on a given commodity each town in a system can expect to obtain from each point in the area served by the system. The probable gross retail income of a town for that good can then be calculated using these proportions together with information on the number of people at each point and per capita expenditure on that good. The repetition of this calculation for different goods and services having different space preference functions would indicate the probable economic support of any town in terms of retail trade. Thus, one value of knowing ratio space preference functions for different commodities is that it is then possible to test the extent to which a spatial supply system is adjusted to spatial demand.

Hence, irrespective of whether individual preference is probabilistic or deterministic, proof of an aggregate ratio preference scale enables predictions of multiple choice behavior, which are impossible with only ordinal information.

Further value in a ratio scale is to be seen in (4.19), for it implies that the probability of choosing one type from a subset of alternatives never observed, can be predicted from the appropriate pairwise probabilities. In other words, the probability of certain unobserved events is predictable.

Finally, the assignment of ratio scores to location types, which represent combinations of town population and town distance, enables the weightings on P and d referred to in (1.1) to be calculated. In fact, only interval scores are necessary to compute such weightings, and therefore for reasons stated above, a multidimensional scaling method is used as described in the next section, to compute these interval scores.

Derivation of an Interval Preference Scale

Given that an interval preference scale can be derived from the observed pairwise proportions, the shape of such a scale should provide interesting information about the exact nature of the trade-off between town population and distance. For the same reason as in deriving the ordinal scale, it is unnecessary to use only independent paired comparisons in deriving the interval scale (see page 96).

The problem is to derive a one-dimensional configuration of points representing location types, such that the distances between points on the scale (s) are a function of the preference dissimilarities between location types as given by $|d| = |p-0.5|$. Now, since each $|\hat{d}|$ is merely as estimate of the true $|d|$, a perfect linear relationship between the $|\hat{d}|$ values and the scale distances (s) cannot be expected.

However, there is a method of obtaining an interval scale arrangement of points with as much accuracy as possible, given the inexact nature of the data. The method, known as Guttman-Lingoes Smallest Space Analysis (SSA-I)[16], is one of many scaling techniques devised to construct metric scales from non-metric data.[17] Whilst the data in this analysis have been shown to have metric properties, the method is equally applicable to metric as to the non-metric data for which is devised. The basic input to the model is a matrix of dissimilarity coefficients which are assumed to indicate only the order of dissimilarity. The algorithm determines a Euclidean configuration of points, in this case location types, in which the distance (s) between each pair of points has the same rank order amongst the inter-point <u>distances</u>, as has that pair's dissimilarity measure ($|\hat{d}|$) amongst the inter-point <u>dissimilarities</u>. In other words the algorithm seeks that configuration of points such that the inter-point distances, s, are some monotonic function of the $|\hat{d}|$ coefficients.

In fact, the method seeks that configuration, which, for the smallest possible sacrifice of monotonicity, requires the fewest

---

[16] Details of this scaling method are provided in Guttman (1968), Lingoes (1965, 1966a, 1966b), Lingoes and Roskam (1970) and Roskam and Lingoes (1970). A more verbal and readable explanation of a similar multidimensional scaling algorithm is given in Kruskal (1964a and b). A general discussion of multidimensional scaling methods, of which SSA-I is but one example, is to be found in Coombs (1964, Chapter 7) and Torgerson (1958, Chapter 11).

[17] Metric data are used here to mean data with interval or ratio scale properties, which give information about how much larger is the distance separating one pair of objects compared to another. By contrast, the strongest nonmetric information obtainable about inter-point separations is which one is the larger. The reason for using a method which requires only nonmetric information to derive a metric scale is provided later in the chapter.

dimensions to locate the points. Obviously, with absolutely no sacrifice of monotonicity it is always possible to represent n points in n-1 dimensions. However, the object of SSA as with all multidimensional routines including factor analysis, is to achieve parsimony in the number of dimensions depicting stimulus and/or subject separations, whilst violating as few as possible of the original coefficients of dissimilarity. Given a parsimonious solution, the test of the veracity of the configuration which is continually applied within the algorithm, is how "closely" the rank order of the initial coefficients can be replicated by the derived scaling of objects. In the case of SSA the closest replication possible is, of course, if the derived scalar distances are some monotonic function of the dissimilarities.

The reason for not seeking a perfect metric relationship in which distances and dissimilarities are related by some fixed formula, as in routines such as metric factor analysis, is that the initial data, if nonmetric, do not justify such an assumption. Also, if nonmetric constraints are imposed in sufficient number, in the form of an ordering of inter-object dissimilarities for <u>many</u> pairs, it can be shown that the nonmetric constraints begin to act like metric ones. Now, if only the rank order of a set of points is known, an ordering of interpoint dissimilarities is possible for only a limited number of pairs. In fact, if A and B are any two points with A lying to the left of B on an ordinal scale, then only for those other pairs of points where one lies to the left of A and the other to the right of B, is it possible to say such pairs have a greater interpoint distance than A and B. With only that limited information, the points on the scale can be moved about quite extensively without violating the inequalities (i.e., without interchanging any two points in the original ranking). However as Shepard (1966, p. 288) indicates:

> as these same points are forced to satisfy more and more inequalities [i.e., statements of which of two pairs of points are more distant] on the interpoint distances ... the spacing tightens up until any but very small perturbations of the points will usually violate one or more of the inequalities.

He goes on to describe the results of a test by Abelson and Tukey (1959, 1963) which indicate how tightly constrained is the metric solution for a set of interpoint inequalities (nonmetric data) as the number of inequalities increases. Essentially they ask what is the smallest possible product-moment correlation between the coordinates which produce a given set of inequalities and the metric coordinate solution derived using these inequalities, which can be achieved without violating monotonicity between the two sets of inequalities. In other words, how big a difference between the given and the computed configuration of points is possible, whilst still satisfying the basic monotonicity constraint? Shepard (1966, pp. 288-289) describes the results for the case of four points on a one-dimensional scale:

> they found that, if only the rank order of the points themselves is known (the ordinal scale), the squared maximin correlation, r^2, is already .65. If, in addition, the rank order of the distances between adjacent points is known (an ordered metric scale), r^2 increases to between .67 and .94 (depending upon the particular ordering given). Finally, if the complete ordering of <u>all</u> interpoint distances is known (a higher ordered metric scale), r^2 increases still further to between .91 and .97 (depending, again, on the particular ordering). For many practical purposes, then, a knowledge of the rank order of the interpoint distances may become almost

> as good as a knowledge of the actual distances themselves.

For the above reasons therefore, it seems unnecessary in this analysis to make strong metric assumptions, particularly since each $|\hat{d}|$ coefficient is not invariant under changes in sample size.

There are several multi-dimensional scaling algorithms with the general characteristics described, which have been developed in the past decade, particularly Shepard's work on the analysis of proximities (Shepard, 1962a, 1962b), Kruskal's MDSCAL (Kruskal, 1964a, 1964b), McGee's HYBRID (McGee, 1966), Young and Torgerson's TORSCA (Young and Torgerson, 1967), Guttman and Lingoes' G-L-SSA-I (Guttman, 1968; Lingoes, 1965, 1966a), and Guttman, Lingoes and Roskam's MINISSA-I (Roskam and Lingoes, 1970). The method used in this analysis is Guttman and Lingoes' Smallest Space Analysis (G-L-SSA-I). It is the most recent working program in the Guttman-Lingoes series for handling the off-diagonal elements of a square, symmetric matrix of coefficients and incorporates recent changes designed to improve the algorithm's ability to minimize loss of perfect monotonicity. In fact it is "equivalent to [the not yet operational] MINISSA-I" (Lingoes, 1970), which Roskam and Lingoes (1970) describe as being:

> Based on extensive empirical studies (Lingoes and Roskam, 1970) of the nonmetric algorithms advanced by Kruskal [Kruskal, 1964a, 1964b], on the one hand, and by Guttman and Lingoes [Guttman, 1968; Lingoes, 1965, 1966], on the other, [and as being] an integrated program containing the best features of both approaches (Roskam, 1969).

Empirical comparisons by the author indicated that G-L-SSA-I is more able to closely approximate perfect monotonicity than MDSCAL, and is therefore a preferable method. No direct comparison has been made between

G-L-SSA-I, TORSCA, HYBRID, and Shepard's method. However, Young and Appelbaum (1968, pp. 22-23) suggest:

> It is probable, although no proof exists that the methods of Lingoes and Torgerson will have less local minima problems [i.e., problems of finding apparently optimal solutions in terms of minimising deviation from perfect monotonicity, when a better solution may exist] than the other methods [Shepard's, Kruskal's and McGee] since the initial configuration is biased in such a way as to be closer to the overall minimum [deviation from perfect monotonicity].

The reason for using G-L-SSA-I rather than TORSCA is that the former is subject to continual revision and improvement, and whilst this is no assurance of its superiority, it would seem to be an advantage. Certainly its authors claim the "virtual elimination of local minimum traps (which are far more frequent than one would suppose without suitable counter-measures)" (Roskam and Lingoes (1970)).

Before describing the scale derived using G-L-SSA-I, the method's goodness-of-fit function is described, which indicates how closely the derived distances between points are a monotonic function of the dissimilarities data. The coefficient of alienation is:

$$\phi = \sum_{j>i}^{n} (s_{ij} - s'_{ij})^2 \; / \; 2 \sum_{j>i}^{n} s_{ij}^2 \qquad (4.20)$$

where s_{ij} = the distance between points i and j as computed by the algorithm, and

s' = the s value with the same rank as $|\hat{d}_{ij}|$ has in the total ranking of $|\hat{d}|$'s (Guttman's rank-images).

The equation is simply half the normalized sum of squared deviations of s_{ij}'s from the ideal. Its minimum is 0 when perfect monotonicity is achieved, and its maximum is 1 when the function can be no less monotonic.

The preference scores of location types, obtained from a one-dimensional scaling of the total sample's matrix of dissimilarity coefficients, are shown in Figure 4.7.[18,19] An interval preference

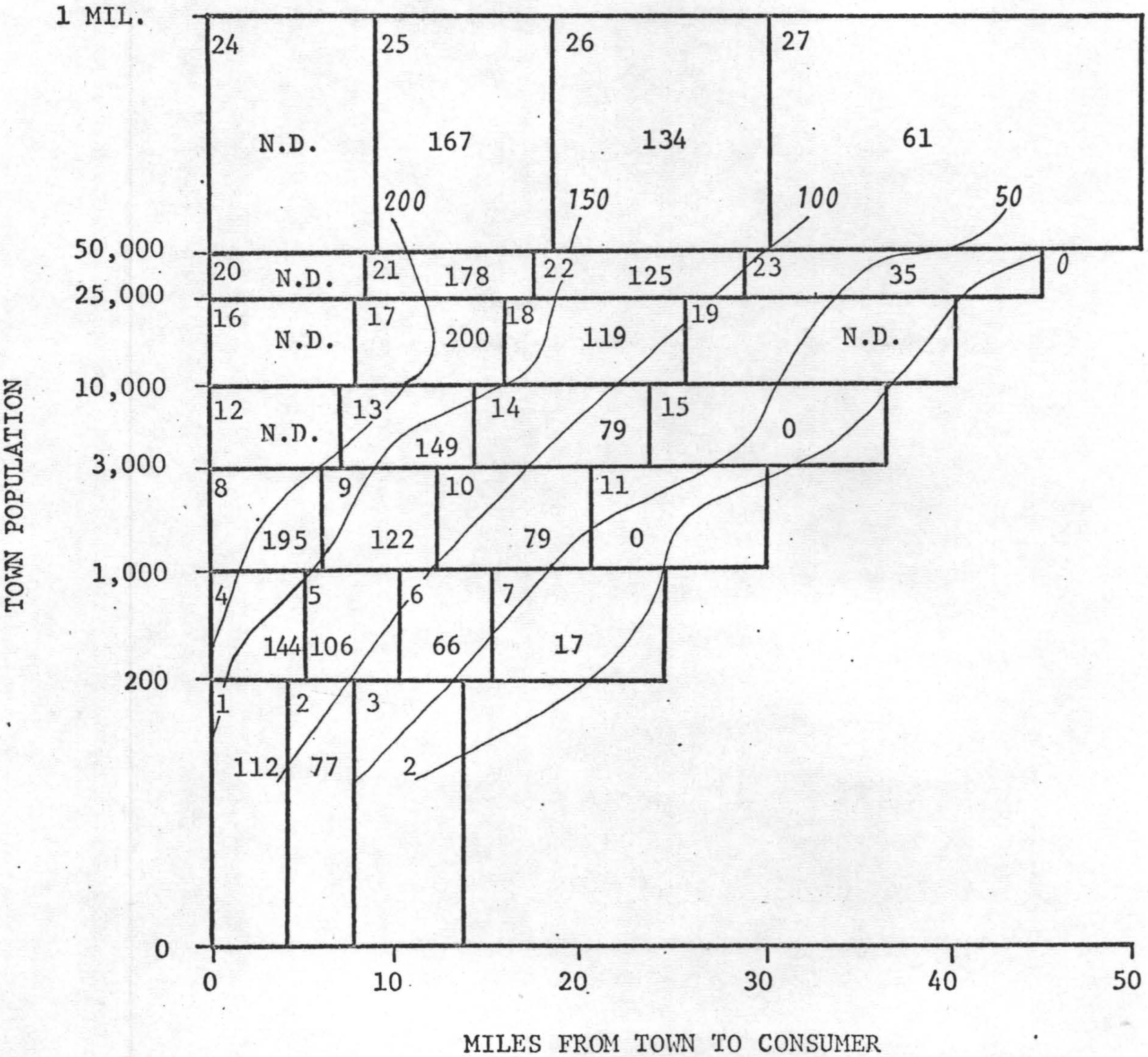
---

[18] The $|\hat{d}|$ matrix is derived from the $\hat{p}$ matrix shown in Table 4.3 with two modifications. The first entails estimating $|\hat{d}|$ for each cell which had no information, using Luce's product rule (equation 4.13), since the SSA-I routine required a value in every cell. The second modification requires the removal from the analysis of all location types whose $|\hat{d}|$ coefficients are invariant over all 26 paired comparisons. Leaving such location types in the analysis results in a degenerate solution. That is, if one location type's $|\hat{d}|$ coefficients invariably equal 0.5, the routine gives a solution where all other location types converge to almost the same point at one end of the scale, and the remaining location type locates at the other extreme. The only invariant $|\hat{d}|$ throughout the study is .5 where one location type is always or never preferred to all others for this sample. Thus it is impossible to scale such a location type metrically, except to say that it must be sufficiently far to the left or right of the location types on the scale to be invariably preferred to or by them.

[19] The coefficient of alienation in the scaling solution is 0.25, which in effect means there is a 25% discrepancy from perfect monotonicity. Considering the varying precision of the dissimilarity estimates, it might be reasonable to attribute some of this coefficient to sampling error. There is, however, no rigorous method of testing this hypothesis. Nevertheless, some indication of the effect of sampling error on the value of the phi coefficient is provided by an experiment by Sherman and Young (1968). Using a two-dimensional set of points in Euclidean space, they randomly perturbed their positions according to a given normal probability distribution and derived a set of inter-point distances from the relocated points. Using TORSCA they then attempted to scale the points in two dimensions and obtained a coefficient of deviation from perfect monotonicity - in this case Kruskal's stress coefficient, which is very similar to the phi coefficient (Kruskal's stress coefficient for the SSA-I analysis of the total sample is .22). The proportion of error introduced into the configuration was defined as $\sigma = x.y$, where y is the standard deviation of the configuration, and x is a constant of proportionality. Four different values of x were chosen, .0, .1, .2, and .5. For 30 points, the stress coefficients corresponding to the above amounts of error are .002, .049, .099 and .221. For 15 points the values were .003, .031, .074 and .173. Thus a considerable amount of the stress value in this analysis might be attributable to sample error.

FIGURE 4.7

Location Types' Interval Preference Scores



Location types 12, 16, 19, 20 and 24 are not scaled using G-L-SSA-I since all are either invariably preferred to or by all other location types with which they are compared. Specifically 19 is invariably preferred by all others, the rest are invariably preferred to all others with which they are compared.

surface is also shown. It is based on linear interpolation of curves on the x axis between the values assigned to the centre of gravity of each location type.[20] Since the y axis is logarithmic the sinuosity of the surface is exaggerated. The trade-off between the utilities of town size and distance is clear. One significant deviation, however, is the bending back of the highest preference curve. This reinforces the pattern noted in indifference curve analyses of the same data (Rushton, 1966; Ewing, 1968). It indicates that large nearby towns, i.e., towns greater than 25,000, are less attractive than only fractionally nearer medium-sized towns (10,000 - 25,000). The implication seems to be that at small distances large towns with associated traffic congestion problems have their size advantage over medium-sized towns negated by congestion. That is, the larger towns become effectively more distant in terms of time than medium-sized towns at a similar physical distance.

The initial hypothesis of this study is that the utility of a central place for shopping is a trade-off function of its size and distance and/or their correlates. Given the interval preference or utility scores and the mean p and d values of each location type as defined by the centre of gravity, it is possible to solve for the functional relationship, $u = f(p,d)$, where u is a utility score.

---

[20]Since each location type has a whole range of p and d values, it is necessary to obtain some single p and d value for each location type, which can be regarded as "most representative" of that type. This is calculated to be the centre of gravity of all alternatives (each with its associated p and d values) available in that location type to any sample member.

The trade-off function is most readily computed with a gravity model of the form:

$$u_{ij} = \frac{p_j^{\beta}}{d_{ij}^{\alpha}} , \qquad (4.21)$$

where u_{ij} = the cardinal utility of alternative j to a consumer at i,

p_j = the population of alternative j,

d_{ij} = the distance from i to j, and

α and β = empirically derived constants.

Equation 4.21 can be rewritten in its logarithmic form as

$$\log u_{ij} = \beta \log p_j - \alpha \log d_{ij} . \qquad (4.22)$$

The equation now has the form of a regression equation and can be solved for the two unknowns, α and β, to determine the average trade-off between population and distance. For the 22 location types scaled, the solution is:

$$\log u = 1.93 + .52 \log p - 1.7 \log d. \qquad (4.23)$$

Rewriting this in gravity model form gives

$$u = 1.93 \times \frac{p^{.52}}{d^{1.7}} . \qquad (4.24)$$

However, since the unit of measurement of u is arbitrary, the constant, 1.93, can be deleted. Thus,

$$u = \frac{p^{.52}}{d^{1.7}} . \qquad (4.25)$$

The multiple correlation coefficient, r, is .756, and therefore the proportion of variance in log u explained by a linear regression on log p and log d is given by the coefficient of multiple determination, r^2, as .579, or 57.9%.

The inference from (4.25) is that the utility of towns increases as the square root of population, i.e., the marginal utility of town population is decreasing. Also, the disutility of distance increases as an exponent (1.7) of distance which means the marginal disutility of distance is an increasing function of distance.

Regarding the specific values of the exponents of p and d in (4.25), a caution is appropriate. In geographical analyses employing gravity models, comparisons are sometimes made between the exponents obtained in different studies. The implication is that different exponents indicate different spatial preferences. This would be a fair inference if it could be shown that each preference surface had constant slope, i.e., if each was a linear and not a curvilinear function of p and d. However, if as in this analysis, the surface is curvilinear, comparison of the gravity equations is inappropriate. Comparison is appropriate only if the ranges of spatial alternatives for which the two surfaces are derived, are the same. Otherwise, differences in two equations' exponents may be simply the result of analysing different, albeit overlapping, parts of the same curvilinear preference surface. An example is provided when those location types with the largest residuals are removed in the present analysis. Specifically, on each of three successive regression analyses the location type in the previous regression with the largest residual was removed. Thus, the third regression analysis lacked not only the originally missing location types 12, 16, 19, 20 and 24, but also 3, 7 and 11. In effect these three least preferred location types of the 22 scaled (see Figure 4.6) tend to increase the negative slope of the surface with respect to d and hence increase d's exponent. As a result of their removal, the exponent drops from 1.7 to 0.85. Thus the shape of the surface derived can be very dependent upon the range of spatial alternatives considered and differences need not reflect preference differences. The implication of the change in exponent for this analysis is that utility declines less sharply at smaller distances. Also, the exclusion of more distant places results in a more linear surface with d's exponent closer to 1, and also results in a higher proportion of explained variation, namely 73.9%, compared to the former 57.9%.

Summary

The following are the major conclusions reached in this chapter. Firstly, a high degree of consistency in households' preferences has been shown. Comparison with the consistency found in hypothetical groups' preferences indicates members of the real sample to have considerable, but not overwhelming, similarity in preferences. Secondly, it has been shown fairly conclusively that the model omits no major explanatory variable. Thirdly, a ratio preference scale has been shown to exist at an aggregative scale. However, it is impossible to determine whether this reflects the same characteristic in individual preference scales. In the Conclusion, a method is discussed which would test for this latter characteristic. Fourthly, the ratio

preference scale has been shown to have more useful properties than the ordinal scale, in terms of:

a) determining the specific weightings on p and d in preference functions;

b) predicting multiple patronage; and

c) predicting probabilities of choice from unobserved sets of alternatives, from a minimum of (n-1) observed pairwise probabilities.

Chapter 5 COMPARATIVE ANALYSIS OF SUBGROUPS' PREFERENCE STRUCTURES

In the previous chapter the implicit assumption was made that variation in subjects' reactions to the same alternatives was attributable solely to differences in the alternatives. Thus,the smaller the difference between alternatives, the more likely are individuals to choose differently between them. However, it is hypothesized that the probability of one alternative being chosen over another is not only a function of their differing amounts of attributes relevant to preference, but also of the differing weights households may attach to the same attributes. The basic goal of the analysis described in this chapter is to determine if households weight the same attributes differently, and if such differences in weighting can be related to identifiable differences in household characteristics or goals.

Methods of Grouping Households for Preference Comparison

In principle, there are two possible approaches to testing this double-barrelled hypothesis. One is to divide the total sample into subsets of households whose town choices and rejections indicate a similarity in their preference rules. The validity of the grouping could be tested by comparing the groups' preference matrices for significant differences. If preference differences were determined, an analysis of household characteristics in each group would determine whether there was any relationship between the characteristics and their preference rules. Unfortunately, this approach is infeasible in this study, due to the inadequate amount of information about preference provided by any single household's choices and rejections of places to patronise. Thus many of the households could have their behavior attributed to such a wide range of preference rules that few households could be said definitely not to share the same preference rule. However, even if adequate data were available, a serious criticism of this kind of approach might be that it is limited by the variety of preference rules which the analyst can conceive, and by his assumptions whether behavior is deterministic or probabilistic. Hence this approach is not used.

The other approach is to divide the total sample into subsets of households, where the groups are univariately or multivariately defined in terms of given amounts of a certain household characteristic. Thus, for example, one subset may consist of all households in the sample with no children, another of households with fewer than three children, and a third consist of all other households. The hypothesis can then be tested whether households with different "amounts of some characteristic" (to be called "variable scores"), also have different preference rules.

Most households are defined in terms of the variables listed in Table 5.1. Whilst it would be possible to group households into many small groups according to their variable scores, and thereby compare the preferences of households with different and similar variable scores, the resultant small sample size of each group would make preference comparison almost impossible. There would be so many empty cells in the preference matrix of one or both groups being compared, that the test would be inconclusive. Thus, the sample is normally divided into only two or three groups, keeping sample sizes in the 100 to 300 range. Details of the "variable score" definition of each group will be provided in the detailed description of the analysis.

Table 5.1

Socio-Economic Attributes, Their Watershed Values

and Associated Group Sample Sizes

Attribute	Watershed value	Sample size	
		Lower group	Upper group
1. Number of persons in household	3	198	258
2. Number of persons 10 years old or less	0	231	225
3. Number of persons between 10 and 20	0	240	216
4. Age of homemaker	43	230	226
5. Homemaker's number of years of formal education	10	161	293
6. Age of farm operator	44	207	246
7. Farm operator's number of years of formal education	10	239	215
8. Number of persons working at off-farm employment	0	261	195
9. Farm acreage	175	217	236
10. Number of years household operated this farm	16	245	211
11. Number of years household lived in this house	13	239	217
12. Net farm income in dollars	2,300	177	219

Improved Estimates of Group Preference Structures

Having defined each group, the accuracy of p estimates can be improved using Luce's method. Using (4.13) and (4.14) for each observed $\hat{p}_{ik}$, other estimates of the same p_{ik} can be computed using $\hat{p}_{ij}$ and $\hat{p}_{jk}$ values, where j refers to each other location type in turn.[1] Rather than average the estimates of each p_{ik}, each estimate, including the observed $\hat{p}_{ik}$, is weighted by its sample size, so as to avoid attributing too much significance to small sample estimates. Also, since two p values are used to estimate p_{ik}, the smaller of the two associated sample sizes is used as the weighting factor. Thus, for all pairs of location types the recalculated value of $\hat{p}_{ik}$ is:

$$\bar{p}^{*}_{ik} = \{ [\sum_{\substack{j=1 \\ j \neq i,k}}^{n} (\frac{\hat{p}_{ij}}{\hat{p}_{ji}} \times \frac{\hat{p}_{jk}}{\hat{p}_{kj}} / (1 + \frac{\hat{p}_{ij}}{\hat{p}_{ji}} \frac{\hat{p}_{jk}}{\hat{p}_{kj}})) \times \mathrm{MIN}(n_{ij}, n_{jk})]$$

$$+ (\hat{p}_{ik} \times n_{ik})\} / \{[\sum_{\substack{j=1 \\ j \neq i,k}}^{n} \mathrm{MIN}(n_{ij}, n_{jk}) + n_{ik}\}, \qquad (5.1)$$

---

[1] There are situations where one or both of the $\hat{p}$'s used to provide an estimate of the third p equal 1 or 0, in which Luce's calculation is impossible. In that case the product rule given in equation 4.13 will contain a zero in the numerator or denominator of the right hand side and so be insoluble. However, if i is invariably preferred to j in the sample and j and k are indifferently preferred or j is preferred to k, it is reasonable to conclude that i will be invariably preferred to k also. In such cases an estimate of p_{ik} using $\hat{p}_{ij}$ and $\hat{p}_{jk}$, is possible without using the product rule, but without violating any of the principles of Luce's method. If, however, i is invariably preferred to j in the sample and k is preferred to j, it is impossible to estimate how much i is preferred to k. For, where $p_{ij} = 1$ is the maximum possible preference measurement, it is impossible to estimate whether the distance separating i and j is infinitely great, or just large enough to avoid a value of p_{ij} less than 1. Thus without a reliable estimate of the distance between members of one pair, a reliable estimate of preferredness between members of the third pair is impossible. Thus situations of the latter type are disregarded in the calculation of $\hat{p}_{ik}$ using $\hat{p}_{ij}$ and $\hat{p}_{jk}$.

where $\hat{p}$ = the observed sample proportion of times one alternative is preferred to another;

$\bar{p}^*$ = the recomputed probability of one alternative being preferred to another; and

$MIN(n_{ij},n_{jk})$ = the minimum of the two sample sizes, n_{ij} and n_{jk}.

An example of the resultant matrix of recomputed preference probabilities for the total sample is shown in Table 5.2 for comparison with Table 4.3. The number of cells with missing data is reduced from 38 to 10 as a result of using the method. Significantly, many of the recomputed values are closer to .5 than the initially observed values. In fact, a comparison of the matrices in Tables 4.3 and 5.2 shows 80 cases where the recomputed $\bar{p}$ is less extreme than the observed $\hat{p}$, and only 27 cases where the opposite is true. This accords well with sampling theory, and with the properties of the binomial distribution, since the smaller the sample size, the more likely are estimates of true proportions to approach the extreme values of 1 or 0.

However, it should be noted that the question of the interdependency of estimates has not been raised. An example will best illustrate the possible interdependencies in estimating a proportionate preference, say $p_{1,3}$. In principle, all the following pairs of estimates can be used to compute an estimate of $p_{1,3}$, although as has been shown, several are likely to be inadmissable for reasons already given and others will be shown to be inadmissable later in the comparison of groups:

$$\begin{array}{cc} p_{1,2}, & p_{3,2} \\ p_{1,4}, & p_{3,4} \\ p_{1,5}, & p_{3,5} \\ \vdots & \vdots \\ p_{1,27}, & p_{3,27} \end{array}$$

Table 5.2

Recomputed Proportionate Preference of Total Sample for Towns in Row Location Types over Towns in Column Location Types

-9.9 SIGNIFIES NO DATA

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	.5																										
2	.3	.5																									
3	.0	.1	.5																								
4	.8	.9	1.0	.5																							
5	.5	.7	1.0	.2	.5																						
6	.1	.3	.8	.0	.1	.5																					
7	.0	.1	.4	.0	.0	.2	.5																				
8	.9	1.0	1.0	.8	.9	1.0	1.0	.5																			
9	.7	.9	1.0	.4	.7	.9	1.0	.1	.5																		
10	.2	.5	.9	.1	.2	.7	.9	.0	.1	.5																	
11	.0	.1	.5	.0	.0	.2	.6	.0	.0	.1	.5																
12	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	.5															
13	.8	1.0	1.0	.7	.8	1.0	1.0	.4	.7	1.0	1.0	0.0	.5														
14	.5	.7	1.0	.2	.5	.9	1.0	.1	.3	.8	1.0	0.0	.2	.5													
15	.1	.2	.8	.0	.1	.4	.8	.0	.0	.3	.8	0.0	.0	.1	.5												
16	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	-9.9	-9.9	1.0	1.0	.5											
17	.7	1.0	1.0	.6	.9	1.0	1.0	.4	.6	1.0	1.0	0.0	.5	.9	1.0	-9.9	.5										
18	.4	.7	1.0	.2	.5	.9	1.0	.0	.2	.7	1.0	0.0	.1	.4	.9	0.0	.1	.5									
19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	.5								
20	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	-9.9	-9.9	1.0	1.0	-9.9	-9.9	1.0	1.0	.5							
21	.8	1.0	1.0	.6	.9	1.0	1.0	.3	.8	1.0	1.0	0.0	.4	.8	1.0	0.0	.4	.9	1.0	-9.9	.5						
22	.5	.8	1.0	.3	.6	.9	1.0	.1	.4	.8	1.0	0.0	.2	.6	.9	0.0	.3	.7	1.0	0.0	.2	.5					
23	.1	.2	.8	.0	.1	.4	.9	.0	.0	.2	.8	0.0	.0	.1	.5	0.0	.0	.1	1.0	0.0	.0	.0	.5				
24	.9	1.0	1.0	.7	.9	1.0	1.0	.4	.9	1.0	1.0	0.0	.5	.9	1.0	-9.9	1.0	1.0	1.0	-9.9	.6	.9	1.0	.5			
25	.7	.9	1.0	.5	.8	1.0	1.0	.3	.6	.9	1.0	0.0	.4	.8	1.0	0.0	.3	.8	1.0	0.0	.4	.7	1.0	.3	.5		
26	.5	.7	1.0	.2	.5	.9	1.0	.1	.3	.8	1.0	0.0	.1	.5	.9	0.0	.0	.5	1.0	0.0	.1	.4	.9	.0	.3	.5	
27	.1	.4	.9	.1	.2	.6	.9	.0	.1	.4	.9	0.0	.0	.2	.6	0.0	.0	.2	1.0	0.0	.0	.1	.7	.0	.1	.1	.5

Inevitably, for example, $\hat{p}_{1,2}$ and $\hat{p}_{1,4}$ are partly interdependent since in some instances where 1 is chosen over 2, it is also chosen over 4. In other instances, however, 1 is chosen over 2 when 4 is absent as an alternative. Thus there is only partial independence in the estimates used to compute $p_{1,3}$. For this reason, the recomputed probabilities shown in Table 5.2 are not used in the Guttman-Lingoes scaling analysis described in Chapter 4. However, it is admissable to use partially interdependent probabilities in comparing groups' preferences, if and only if the number of statistically significant differences between random groups' partially interdependent probabilities is used as a benchmark to judge the non-random groups' number of significant differences.[2] Since the original proportions shown in Table 4.3 are also partially interdependent, such a comparison with random group results would be necessary in any case. Thus, the use of recomputed probabilities seems justified for two reasons. Firstly the disutility of partially interdependent probabilities is compensated for by using the amount of "significant" difference between random groups as a benchmark to judge the amount of difference between non-random groups. Secondly, added information is gained from the independent component of pairs of $\hat{p}$'s.

The Selection of a Method for Comparing Groups' Preference Structures

Having defined each group and derived its preference matrix using $\bar{p}^*$ rather than $\hat{p}$ estimates, a method is required to compare pairs of preference matrices for significant differences. Several methods might seem

[2]The details of this comparison between random and non-random groups' results are described later in the chapter.

appropriate. For example, it is possible to compare two groups' ranking of location types. For several reasons, however, it would be difficult to place confidence in the statistical reliability of the result. Firstly, rankings are not invariant under changes in the ranking algorithm used. At least four such algorithms were discussed in Chapter 4, and more could be devised. Certainly all do not produce the same ranking and presumably the differences between two groups' rankings might vary between algorithms. However, even assuming this error source were sufficiently small to disregard, none of the ranking algorithms allows for the sampling error associated with each estimate of p. Thus, any ranking obtained is not invariant under changes in the sample size associated with each $\hat{p}$. Therefore, it is difficult to know the confidence to place in a comparison of rankings which are subject to variation due to sample size.

This latter argument can also be used against comparing G-L-SSA-I scale values of location types derived from the dissimilarity matrices of two groups, since estimates of d are subject to the same error as the estimates of p from which they are derived. Thus, the scalar position of a location type is not invariant under changes in sample size. In addition, since the scale derived by G-L-SSA-I is an interval and not a ratio scale, the relationship between points on the scale is not altered by any uniform stretching of the scale. In every case one location type always lies at one extreme of the scale (-100), and one at the other (+100), irrespective of the proportion of times the one is preferred to the other. Thus, it is unreliable to infer differences in the amount two groups' prefer one thing to another, from the distances the two things are apart on the two scales. An example of the differences between the scale values of two groups of similar size drawn randomly from the total sample is provided in Figure 5.1.

The third of the methods for testing for preference differences, and the one used here, is to statistically test for significant differences in the two groups' estimates of p_{ij} for all pairs of p estimates in the matrices. Depending on how often the two groups significantly differently prefer one location type to others, that location type can be defined as more preferred by one group than the other. Repeating this for all 27 location types, and provided sample sizes associated with the tests are not small, preference differences between the two groups can be identified.[3]

In each comparison of two groups' p_{ij} estimates, the null hypothesis is:

$$H_0: \; p_{ij1} = p_{ij2}$$

and the alternative hypothesis is:

$$H_1: \; p_{ij1} \neq p_{ij2}$$

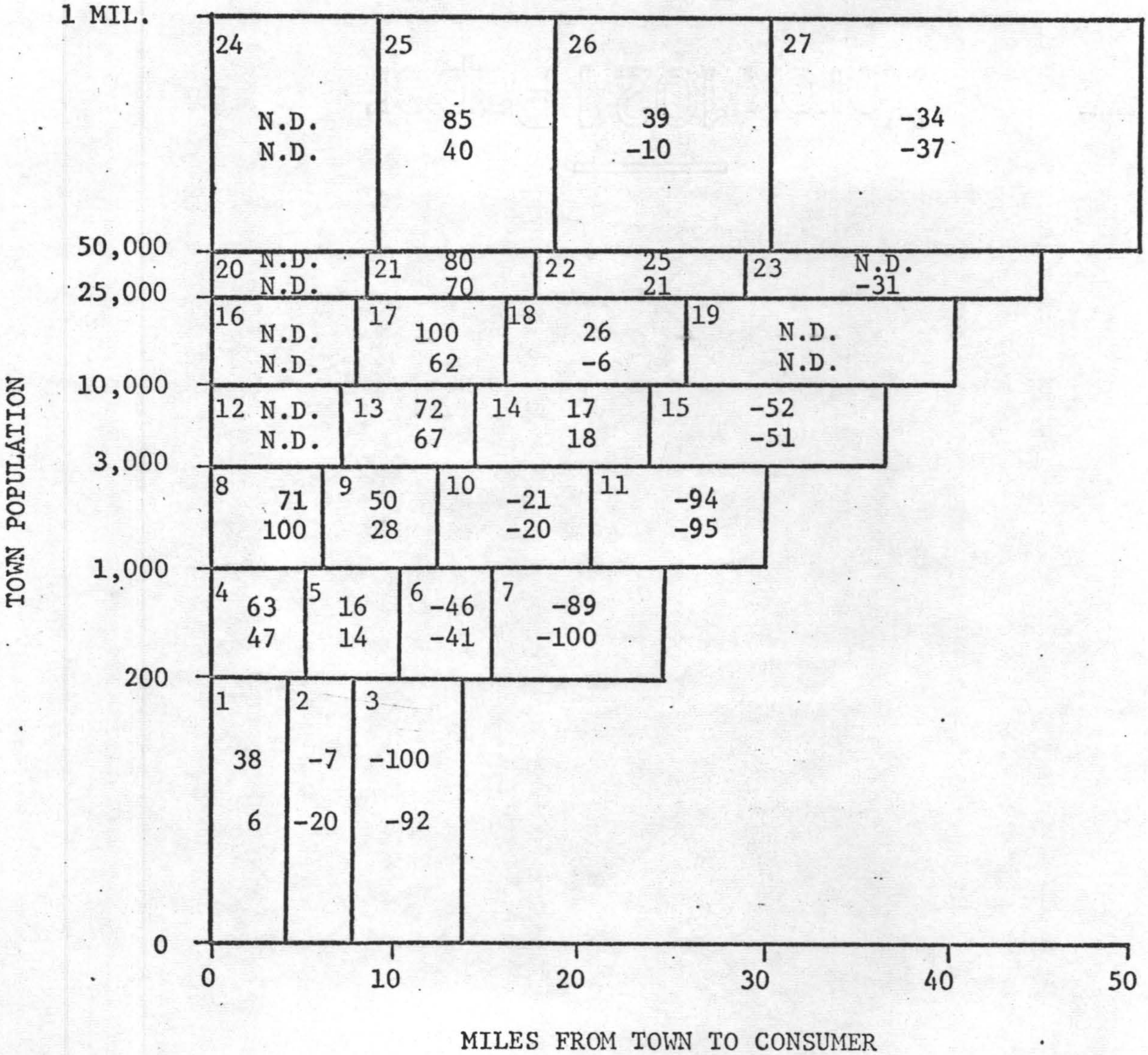
where p_{ijk} = the true proportion of times the total population represented by k prefers alternative i to j. To accept the alternative hypothesis requires that the difference between the two $\bar{p}^*_{ij}$'s be greater than might <u>commonly</u> be expected between two random samples of the <u>same</u> p_{ij}.

---

[3] If sample size is very small, e.g., less than 5, it requires very large differences between estimates of p for a statistically significant difference to be inferred. The use of Luce's product rule to compute $\bar{p}^*$, however, normally provides adequate sample sizes to test for differences.

Figure 5.1

A Comparison of the Guttman-Lingoes SSA-I Scale Values Derived from the Dissimilarity Matrices of Two Independent Groups Drawn Randomly from the Total Sample



Upper group's sample size = 215

Lower group's sample size = 240

In this case a 5% chance of a type I error is taken, i.e., $\alpha = 0.05$.

In Chapter 4 it was assumed and subsequently verified that $P\{x|T\}$ is constant for all trials in the sample, and by implication that $P\{x|x \cup y\}$ is constant. However, it is arguable that such verification may not necessarily be absolute but rather imply virtual constancy of probabilities. In other words, the test may have been insensitive to minor variations in probabilities. Hence, it is hypothesized that there are cases where small but significant differences between certain pairwise probabilities of groups exist. Thus, it is assumed that $P\{x|x \cup y\}$ is constant for all trials within defined subgroups of the sample.

Since the p_{ij} estimates are based on binomial random variables, it is unnecessary to generate a random sampling distribution of $\bar{p}^*_{ij}$ differences against which to compare observed inter-group differences. Instead, observed differences can be compared with theoretical random sampling distributions as a means of testing whether the observed differences are larger than might normally occur by chance. First, the statistics used for the test on each pair of $\bar{p}^*$'s will be described, followed by a definition of how often two groups must prefer one location type significantly differently to others, to regard the groups as significantly differently preferring that location type.

Whilst the same statistic can be used in comparing all p estimates, it is computationally more expedient to use one of two statistics on any pair, depending upon the associated sample sizes. The sample size requirements of the first statistical test, the χ^2 test for two independent samples, are best described by an example. Members of each group can be classified only two ways, either as choosing one alternative or the other. Thus, a two-by-two contingency table is sufficient to contain every possible event (see Figure 5.2). Each cell describes the number of times one group chooses one alternative.

Figure 5.2

	Group		
Alternative	1	2	Total
i	A	B	A+B
j	C	D	C+D
Total	A+C	B+D	N

$$A = n^*_{ij1}\, \bar{p}^*_{ij1}$$

$$C = n^*_{ij1}(1 - \bar{p}^*_{ij1})$$

$$B = n^*_{ij2}\, \bar{p}^*_{ij2}$$

$$D = n^*_{ij2}(1 - \bar{p}^*_{ij2})$$

$$N = n^*_{ij1} + n^*_{ij2}$$

$$n^*_{ij} = n_{ij} + \sum_{\substack{k=1 \\ k \neq i,j}}^{n} \mathrm{MIN}(n_{ik}, n_{jk})$$

To use the χ^2 test requires that either N exceed 40, or if N is between 20 and 40, that all expected frequencies be greater than or equal to 5. The expected frequency in any cell is obtained by multiplying the two marginal totals common to that cell and dividing this product by N. Thus

$$A_{EXP} = \frac{(A + B) \times (A + C)}{N} \tag{5.2}$$

where A_{EXP} = the expected number of times alternative i is chosen by group 1. The null hypothesis that there is no significant difference between the groups' $\bar{p}^*_{ij}$ may be tested by:

$$\chi^2 = \sum_{m=1}^{2} \sum_{n=1}^{2} \frac{(O_{mn} - E_{mn})^2}{E_{mn}} \tag{5.3}$$

where O_{mn} = the observed number of cases in the mth row of the nth column, and

E_{mn} = the number of cases expected under the null hypothesis in the mth row of the nth column. This equation covers the general case of comparing two independent samples with two possible responses. The χ^2 values yielded by the formula are distributed approximately as chi square with 1 degree of freedom. The sampling distribution of χ^2 shows the probabilities, given the null hypothesis is true, of various χ^2 values. If the probability of a computed value of χ^2, together with the probabilities of obtaining any larger value of χ^2, is relatively small under the null hypothesis, say 0.05, then it can be said with 95% certainty that such a χ^2 value indicates the two samples are drawn from different populations.

In this analysis, given 1 degree of freedom and $\alpha = 0.05$, the null hypothesis is rejected if the computed χ^2 is greater than 3.84.

Where only two responses are possible for each group, however, Siegel (1956, p. 107) suggests the following derivation of χ^2, since it "incorporates a correction for continuity which markedly improves the approximation of the distribution of the computed χ^2 by the chi-square distribution":

$$\chi^2 = \frac{N(|AD - DC| - [N/2])^2}{(A+B)(C+D)(A+C)(B+D)} \tag{5.4}$$

with one degree of freedom.

The second test, the Fisher exact probability test is used only if conditions for the χ^2 test are not met. Its particular advantage is that it can handle very small samples, given that they can be described in a 2 × 2 contingency table, such as in Figure 5.1. Given that the marginal values are fixed, the test computes the probability of the actual cell frequencies or any more extreme frequency distribution. Thus, if column 1 has values 7 and 2 and column 2 has values 3 and 6 for alternatives i and j respectively, more extreme frequency distributions would be 8 and 1, 2 and 7 for groups 1 and 2 respectively, and 9 and 0 and 1 and 8.

The exact probability of observing a particular set of frequencies in a 2 × 2 table, when the marginal totals are fixed, is given by the hypergeometric distribution:

$$P(A,B,C,D) = \frac{\binom{A+C}{A}\binom{B+D}{B}}{\binom{N}{A+B}} \tag{5.5}$$

which can be rewritten as

$$P(A,B,C,D) = \frac{(A+B)!(C+D)!(A+C)!(B+D)!}{N!A!B!C!D!} . \tag{5.6}$$

The Fisher exact probability test entails computing this probability for the given frequency distribution and all more extreme distributions. If the probability is less than or equal to 0.05, the null hypothesis that the two samples are drawn from the same population is rejected. As in the χ^2 test, therefore, $\alpha = 0.05$.

However, Siegel (1956, p. 102) describes a slight modification of the Fisher test by Tocher (1950) which "provides the most powerful one-tailed test for data in a 2 × 2 table". The modification is designed to cover situations where the sum of probabilities of both the observed frequency and all more extreme frequencies would result in accepting the null hypothesis (H_o) of no difference between groups, but where the sum of probabilities associated with all the more extreme frequencies, and not that observed, would indicate rejection of H_o. The modification in effect tries to answer the question as to how much importance should be attributed to the actual frequency distribution observed, since repeated samples from the same population are likely to vary. Tocher suggests in cases where the smaller probability indicates H_o should be rejected and the larger indicates it should be accepted, that the following ratio be computed:

$$\frac{\alpha - \text{Pr(more extreme cases)}}{\text{Pr(observed case taken alone)}} \tag{5.7}$$

where α = probability of a type I error. If a random variable drawn from a uniform distribution between 0 and 1, is less than the ratio, H_o should be rejected, but if larger, H_0 is accepted. Thus, the closer the second term in the numerator approaches α, the closer the ratio approaches 0 and therefore the smaller is the chance of the random variable being less than the ratio. In effect the modification implies that in a certain proportion of cases where the sum of probabilities exceeds α it is still justifiable to reject the null hypothesis, due to the likely variability in such frequencies over repeated samples. Just what proportion is acceptable is, of course, defined by the ratio in (5.7), and since the frequencies are subject to random error, it is necessary to employ a randomization procedure to determine if this particular case should be regarded as indicating significant group differences or not.

Earlier the need was indicated for a criterion to judge whether the preference structures of two groups were significantly different. In principle, if each p_{ij} estimate were independent, and $\alpha = 0.05$, more than 5% of two groups' $\bar{p}^*_{ij}$'s would require to be statistically different in order to consider the preference structures significantly different. However, the p_{ij} estimates in each group are only partially independent and therefore the 5% criterion is inappropriate. Instead the method used to judge whether a given percentage of significant differences could be attributable to random differences is to compare the preferences of random samples drawn from the total sample. Each household in the total sample is randomly assigned to one of two groups and the proportion of "significant" differences between the two groups' p estimates computed. The process is repeated 50 times to provide a distribution of significance percentages (see Figure 5.3).[4] The mean of the distribution is 5.8%. Whilst the

[4] Ideally, each preference matrix contains 351 paired comparisons and

theoretical 5% significant difference is not used as a reference, it is notable that the average percentage derived from random groups with only partially independent p_{ij} estimates is not much greater than 5%.[5]

---

therefore 351 possible sources of significant difference. However, certain cells can be ruled out as possible sources of difference, either because in one or both matrices there is no information about preference between a pair of location types, or because the total sample's preference proportion for that pair equals 1 or 0, and is therefore invariant for any subset of the sample. Thus, in each of the 50 iterations, certain of the 351 cells are eliminated from the computation of the significance percentage.

[5] Regarding the small difference, it should be noted that each group's preference matrix is not computed independently using Luce's product rule. This was the approach used initially, but the average percentage of "significant" differences per iteration was 15%, indicating a strong bias has been introduced in calculating the $\bar{p}^*$ values. The source of the bias was the algorithm (an earlier version of RECOMP in Appendix B) used to compute $\bar{p}^*_{ik}$, described in footnote 1 in this chapter. Treating each group independently often resulted in different $\hat{p}_{ij}$'s and $\hat{p}_{jk}$'s being excluded from the calculation of each group's $\bar{p}^*_{ik}$. Thus, for example the table below shows characteristic sets of $\hat{p}_{ij}$ and $\hat{p}_{jk}$ values originally used to calculate $\bar{p}^*_{ik}$ for two groups:

Group 1

$$\bar{p}^*_{1,2} = f\begin{bmatrix} \hat{p}_{1,3}, \hat{p}_{2,3} \\ \hat{p}_{1,4}, \hat{p}_{2,4} \\ \hat{p}_{1,5}, \hat{p}_{2,5} \\ \hat{p}_{1,6}, \hat{p}_{2,6} \end{bmatrix}$$

Group 2

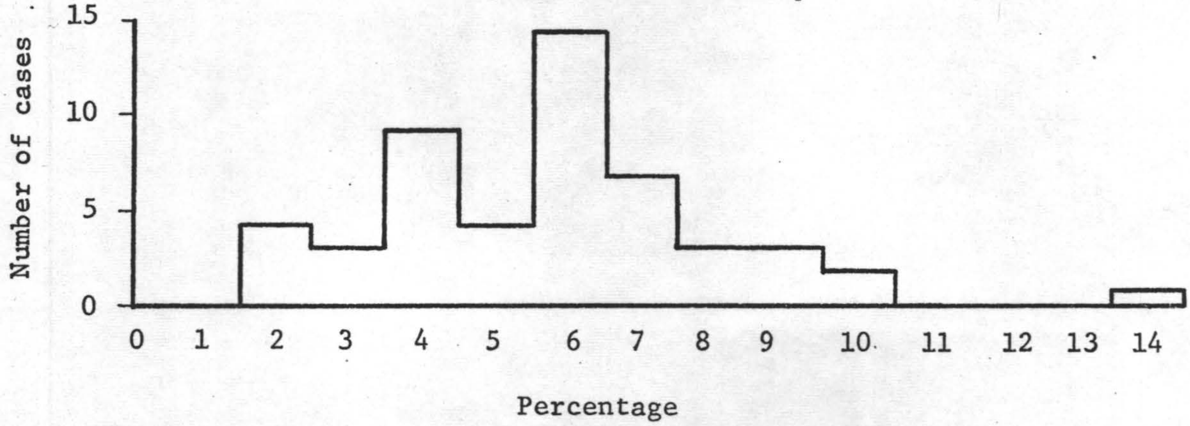
$$\bar{p}^*_{1,2} = f\begin{bmatrix} \hat{p}_{1,6}, \hat{p}_{2,6} \\ \hat{p}_{1,7}, \hat{p}_{2,7} \\ \hat{p}_{1,8}, \hat{p}_{2,8} \\ \hat{p}_{1,9}, \hat{p}_{2,9} \end{bmatrix}$$

The calculation of the two $\bar{p}^*_{1,2}$ values therefore showed biases dependent on the different $\hat{p}$ values used to calculate each. In addition, such biases would often be repeated in calculating $\bar{p}^*$ values containing common location types, e.g., $\bar{p}^*_{1,2}, \bar{p}^*_{1,3}, \ldots, \bar{p}^*_{1,27}$. This "bandwagon" effect is illustrated below.

$$\bar{p}^*_{1,2} = f\begin{bmatrix} \hat{p}_{1,3}, \hat{p}_{2,3} \\ \hat{p}_{1,4}, \hat{p}_{2,4} \\ \vdots \quad \vdots \\ \hat{p}_{1,6}, \hat{p}_{2,6} \end{bmatrix}, \bar{p}^*_{1,3} = f\begin{bmatrix} \hat{p}_{1,2}, \hat{p}_{3,2} \\ \hat{p}_{1,4}, \hat{p}_{3,4} \\ \vdots \quad \vdots \\ \hat{p}_{1,6}, \hat{p}_{3,6} \end{bmatrix}, \ldots, \bar{p}^*_{1,5} = f\begin{bmatrix} \hat{p}_{1,2}, \hat{p}_{5,2} \\ \hat{p}_{1,3}, \hat{p}_{5,3} \\ \vdots \quad \vdots \\ \hat{p}_{1,5}, \hat{p}_{5,6} \end{bmatrix}$$

Figure 5.3

Histogram of the Percentage of "Significantly" Different Paired Comparisons in the Preference Structures of 50 Pairs of Independent Random Groups



The above distribution for random subgroups of the total sample provides a benchmark, against which to evaluate the true significance of observed amounts of "significant" differences between non-random groups. The general hypothesis tested here for each of 12 household attributes is that households with lower scores on an attribute have significantly different locational preferences than households with higher scores. Thus, for example, it is hypothesized that wealthier households have differing preferences than less wealthy ones, etc. Attributes of the type listed in Table 5.1 are often alluded to as possible sources of differing spatial preferences. Whilst there are others not available to this analysis which could be hypothesized to have a bearing on preference structures, those available provide an adequate cross-section of household variables conventionally regarded as "explainers" of behavioral variation.

Preference Comparisons of Twelve Pairs of Social and Economic Groups

For each attribute the total sample is partitioned into two groups characterized respectively by lower and higher scores on the attribute. Since for most of the variables, there are no established hypotheses as to critical watersheds in the relationship of these variables to behavior,

---

Clearly, in the example similar sets of p_{1j} are used in the calculation of different $\bar{p}^*_{1k}$'s. This means that if two groups' $\bar{p}^*_{1k}$'s values are "significantly" different for one value of k, they are likely to be so for several values of k, since many of the p_{1j}'s used in the calculation are common to several $\bar{p}^*_{1k}$'s for one group.

The problem is overcome by calculating the groups' $\bar{p}^*_{ik}$'s concurrently. Thus, in calculating $\bar{p}^*_{ij}$, only those $\hat{p}_{ij}$'s and $\hat{p}_{jk}$'s are used which meet the required conditions to use Luce's product rule in both groups. The resultant large decline in "significant" difference percentages would seem to indicate that this source of bias is eliminated.

the choice of a watershed value for each attribute is relatively arbitrary. Thus,initially values are chosen which maximize the sample size of both groups. If small samples were used to determine groups' preference structures, any apparent absence of preference differences might simply be a function of the small sample size, since the smaller the sample size, the larger must be the difference between two p estimates in order to reject the null hypothesis with reasonable confidence. Table 5.1 shows the watershed value for each variable and the sample size of each group. Due to incomplete information particularly on the attributes age, education and income, the total sample size associated with each grouping varies. Only with respect to income is the sample size any less than 450, specifically 396.

The percentage of pairwise preferences which are significantly different for each pair of groups is shown in Table 5.3. It is clear

Table 5.3

Percentages of Significantly Different Paired Comparisons in Socio-Economic Groups' Preference Structures

Attribute	Watershed value	Percentage of significant differences
1. Number of persons in household	3	6
2. Number of persons 10 years old or less	0	9
3. Number of persons between 10 and 20	0	9
4. Age of homemaker	43	6
5. Homemaker's number of years of formal education	10	8
6. Age of farm operator	44	10
7. Farm operator's number of years of formal education	10	4
8. Number of persons working at off-farm employment	0	14
9. Farm acreage	175	4
10. Number of years household operated this farm	16	9
11. Number of years household lived in this house	13	10
12. Net farm income in dollars	2,300	7

from the percentages, as compared to the distribution of percentages for random groups in Figure 5.3, that only one pair of groups can be regarded as having significantly different preferences, namely the group with nobody employed off the farm and the group with one or more household members employed off the farm.

In order to determine the specific location types which the two groups prefer significantly differently, it is necessary to compare the proportion of times the two groups prefer each location type significantly differently (see Table 5.4) with the proportion of times

Table 5.4

The Percentage of Times Households with Nobody Working Off-Farm Prefer each Location Type Significantly Differently from Households with One or More Members Working Off-Farm.

Location type	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Percentage of significant differences	9	32	0	14	23	29	0	0	14	36	14	0	9	18	14

Location type	16	17	18	19	20	21	22	23	24	25	26	27
Percentage of significant differences	0	5	64	0	0	10	5	11	0	0	5	5

two random groups would prefer the same location type significantly differently. The latter proportion for two random groups is shown in Table 5.5 for each of the 27 location types and for 50 different pairs of random groups drawn from the total sample. This provides 50 random percentages for each location type against which to compare the non-random percentages in Table 5.4. As might be expected, some location types are more likely to be "significantly" differently preferred than

Table 5.5

The Proportion of Statistically Significant Preference Differences for Each Location Type for 50 Pairs of Independent Random Groups

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	.0	0.00	0.00	0.00	[illegible]	0.00	.09	0.00	0.00	0.00	.05	.09	0.00	0.00	0.00	0.00	0.00	.18	0.00	0.00	.05	.18	0.00	0.00	0.00
2	.5	.32	.24	.36	.14	.15	.14	.05	.05	.14	.14	.18	.09	.09	.09	.14	.09	.50	.23	.14	.23	.14	.09	.14	.27
3	.10	.14	0.00	0.00	[illegible]	0.00	.05	.05	0.00	0.00	0.00	0.00	.05	.05	.14	.05	.05	.05	0.00	.05	0.00	.09	.05	0.00	.14
4	.0	0.00	.05	.05	[illegible]	.05	0.00	0.00	0.00	.05	0.00	0.00	.05	0.00	.09	0.00	.05	.18	.09	.18	.14	0.00	.27	.09	.09
5	.14	.09	.05	.09	.05	.15	.05	0.00	.14	.09	.09	.05	.18	.05	.09	.09	.23	.14	.14	.32	.14	.05	.14	.14	.14
6	.5	.14	.05	.05	[illegible]	.05	.32	0.00	0.00	.09	.14	.09	.27	0.00	.05	.05	.18	.14	.23	.14	.05	.18	.19	.05	.09
7	.5	.05	0.00	0.00	0.00	0.00	.09	0.00	0.00	.05	0.00	.05	0.00	.05	0.00	.05	0.00	.10	.05	.05	0.00	0.00	0.00	0.00	0.00
8	.0	.05	0.00	.05	0.00	0.00	0.00	0.00	.10	0.00	0.00	0.00	0.00	0.00	0.00	.05	0.00	0.00	.05	0.00	0.00	0.00	0.00	0.00	0.00
9	.9	.05	0.00	.05	.09	.10	.09	.09	.05	.10	0.00	.05	.09	0.00	.05	.09	.14	.09	.18	.18	.14	0.00	.18	.09	.18
10	.3	.15	0.00	.05	[illegible]	.05	.18	.09	0.00	.05	.18	.09	.14	.10	.27	0.00	.14	.18	.05	.09	.09	.32	.05	.09	.05
11	.0	.10	0.00	.05	[illegible]	.05	.18	0.00	0.00	.05	.05	.05	.10	0.00	0.00	.05	.05	.10	.05	.18	0.00	.05	0.00	.05	.05
12	.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	.14	0.00	0.00	0.00	0.00	0.00	0.00	.05	0.00	0.00	.05	.05	0.00	0.00	.05	.05	.10	.09	.09	.14	.14	0.00	.05	.14	0.00
14	.10	.05	.05	.05	.14	.10	.05	.05	.05	.05	.18	0.00	.09	.05	.05	.18	0.00	.36	.14	.18	0.00	.05	.09	.09	.05
15	.5	.18	.05	.09	.05	.05	.18	0.00	.05	.45	.23	.18	.27	.14	.05	.18	.14	.23	.05	.14	0.00	.05	.14	.27	.19
16	.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17	.0	0.00	0.00	0.00	[illegible]	.15	0.00	0.00	0.00	0.00	0.00	0.00	.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	.09	0.00	0.00	0.00
18	.5	0.00	.14	.05	.32	0.00	.14	0.00	0.00	0.00	.05	0.00	0.00	0.00	0.00	0.00	0.00	.05	.09	0.00	0.00	.18	0.00	0.00	.32
19	.0	0.00	0.00	0.00	[illegible]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20	.0	0.00	0.00	0.00	[illegible]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21	.0	0.00	.05	0.00	[illegible]	0.00	.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	.05	0.00	.10	.05
22	.0	.09	.05	.05	.32	.05	0.00	0.00	0.00	.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	.27	0.00	0.00	.09	.14	0.00	0.00	.09
23	.5	.05	0.00	0.00	.09	0.00	0.00	0.00	0.00	.10	.09	.17	0.00	0.00	0.00	.11	0.00	.09	0.00	0.00	0.00	.15	0.00	0.00	.11
24	.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25	.0	.05	0.00	0.00	[illegible]	0.00	0.00	0.00	.05	.05	0.00	0.00	0.00	0.00	0.00	0.00	.05	.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00
26	.0	0.00	.14	0.00	.35	0.00	0.00	.05	0.00	0.00	0.00	0.00	0.00	.10	0.00	0.00	.05	.27	.05	0.00	.05	.05	.05	.05	.14
27	.0	.05	0.00	.05	0.00	.05	.05	.05	0.00	0.00	.05	0.00	0.00	.05	0.00	.05	.05	.05	0.00	.14	0.00	0.00	0.00	0.00	0.00
Column mean	.6	.07	[illegible]	.04	.07	.03	.07	.02	.02	.06	.06	.04	.06	.03	.04	.05	.06	.14	.06	.08	.05	.08	.06	.06	.09

Table 5.5 (cont'd.)

	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	.0	.09	.00	.24	.05	.05	0.00	0.00	0.00	0.00	0.00	0.00	.05	0.00	.05	0.00	.05	.14	.59	.05	0.00	.14	0.00	0.00	.09
2	.14	.18	.14	.14	.1	.18	.27	.18	.05	.18	.18	.14	.09	0.00	.18	.18	.14	.09	.09	.23	.14	.09	.09	.05	.32
3	.5	.10	.05	0.00	.07	.14	.14	.14	0.00	.14	0.00	.05	.05	0.00	0.00	0.00	.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	.14	0.00	0.00	.10	0.00	.18	0.00	.09	0.00	.05	0.00	0.00	.05	.05	.09	.05	.09	.09	.09	0.00	0.00	0.00	.14	.05	0.00
5	.9	.09	.09	.38	.14	.05	.14	.18	.05	.14	0.00	.05	0.00	0.00	.05	.05	.05	.14	.23	0.00	.05	.10	.18	.05	.05
6	.14	.23	.14	.05	.1	.05	.14	.18	0.00	.18	.14	.19	0.00	0.00	.23	.09	0.00	.05	.09	.09	.10	.05	.23	0.00	.14
7	.0	0.00	.05	0.00	.0	0.00	.15	0.00	0.00	.14	.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	.05	0.00	0.00	.10	.05	.05	0.00
8	.0	0.00	0.00	0.00	0.0	.05	0.00	.05	0.00	0.00	.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	.05
9	.9	.18	.05	.10	.05	.05	.09	.14	0.00	.23	0.00	.05	0.00	0.00	.09	.19	.09	.18	.09	0.00	.05	.05	.05	.09	0.00
10	.1	.05	.14	.14	.1	.18	.18	.09	.09	.29	.09	.29	.05	.05	.09	.14	.09	.09	.09	.18	.10	0.00	.09	.36	0.00
11	.0	.10	0.00	.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	.05	.14	.05	.10	0.00	.09	.14	0.00	0.00	.05	0.00	.14
12	.0	0.00	0.00	0.00	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	.0	.05	0.00	.05	0.0	.09	.05	.14	0.00	.05	.05	0.00	0.00	0.00	.05	0.00	0.00	.05	0.00	0.00	0.00	0.00	.05	.05	0.00
14	.4	.05	.09	.10	.05	.27	.14	.50	.23	.14	.09	.10	0.00	.05	.05	.05	.18	.23	.14	.18	.24	0.00	.09	.09	.05
15	.14	.14	.14	.14	.07	.10	.14	.14	.14	.19	.14	.10	.14	0.00	.23	.18	.05	.27	.36	.09	.10	.09	.32	.05	.09
16	.0	0.00	0.00	0.00	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17	.0	0.00	.05	0.00	0.0	.05	0.00	0.00	.09	.05	0.00	0.00	0.00	.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18	.5	.18	.05	0.00	0.0	0.00	.05	.05	.05	.41	0.00	0.00	0.00	0.00	.09	0.00	.05	.14	.05	.09	0.00	.14	.05	0.00	0.00
19	.0	0.00	0.00	0.00	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20	.0	0.00	0.00	0.00	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21	.0	0.00	.05	.05	0.0	0.00	0.00	0.00	.05	0.00	0.00	0.00	.05	.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
22	.5	0.00	0.00	.05	0.0	.23	0.00	.05	.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	.05	.27	.05	.05	0.00	0.00	.05	0.00	.05
23	.5	0.00	.05	0.00	.1	.10	0.00	0.00	0.00	.06	0.00	0.00	0.00	0.00	0.00	0.00	.11	0.00	.09	0.00	0.00	.11	0.00	0.00	0.00
24	.0	0.00	0.00	0.00	0.0	0.00	0.00	0.00	.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25	.0	0.00	0.00	0.00	0.0	.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	.05	.05	0.00	.05
26	.0	.09	0.00	0.00	.27	.05	0.00	.14	0.00	.05	0.00	0.00	0.00	.05	0.00	.05	.05	.09	.05	.18	0.00	.05	0.00	0.00	0.00
27	.18	.05	.23	.05	0.00	.09	0.00	.05	0.00	.05	.09	0.00	0.00	.05	.05	0.00	.53	.09	.05	0.00	0.00	0.00	0.00	0.00	.09
Column mean	.7	.07	.06	.07	.0	.05	.06	.09	.04	.10	.04	.04	.02	.02	.06	.05	.07	.08	.10	.06	.03	.04	.06	.04	.05

others (see Table 5.5). For example, a location type such as 17 is rarely "significantly" differently preferred, whereas type 10 is more frequently so. Inevitably where some location types are almost invariably preferred to all others, the chance of a significant difference in group preferences is reduced. Thus, for location type 17, a percentage of significantly different preferences greater than 9% occurs only once in 50 cases, whilst for location type 10 the proportion exceeds 9% in 22 of the 50 random groupings. As a result a different critical value is used for each location type in determining whether a particular percentage of significant preference differences is larger than might be attributable to chance. Given the 50 random values for each location type, only proportions greater than 94% of the random values, i.e., greater than 47 of the 50 values, are regarded as unlikely to be due to chance. This involves a 6% chance of a type I error. Table 5.6 shows for each location type the four highest proportions of significant differences in the 50 random groupings. To be considered significant, a proportion should at least exceed the lowest of the four values, i.e., the 47th largest of the 50 values.

Figure 5.4 shows those location types satisfying the above condition for the grouping related to a household's number of off-farm workers. Inspection of the statistical tests relating to each of those location types reveals the direction of the preference differences for each location type (see Figure 5.4). Interpreting the pattern, inferences can be made both about the location types which show up as significantly differently preferred and those which do not. There are three possible explanations for absences of significant differences. With regard to certain location types such as 12, 16, 20 and 19, the total sample always chooses the first

Table 5.6

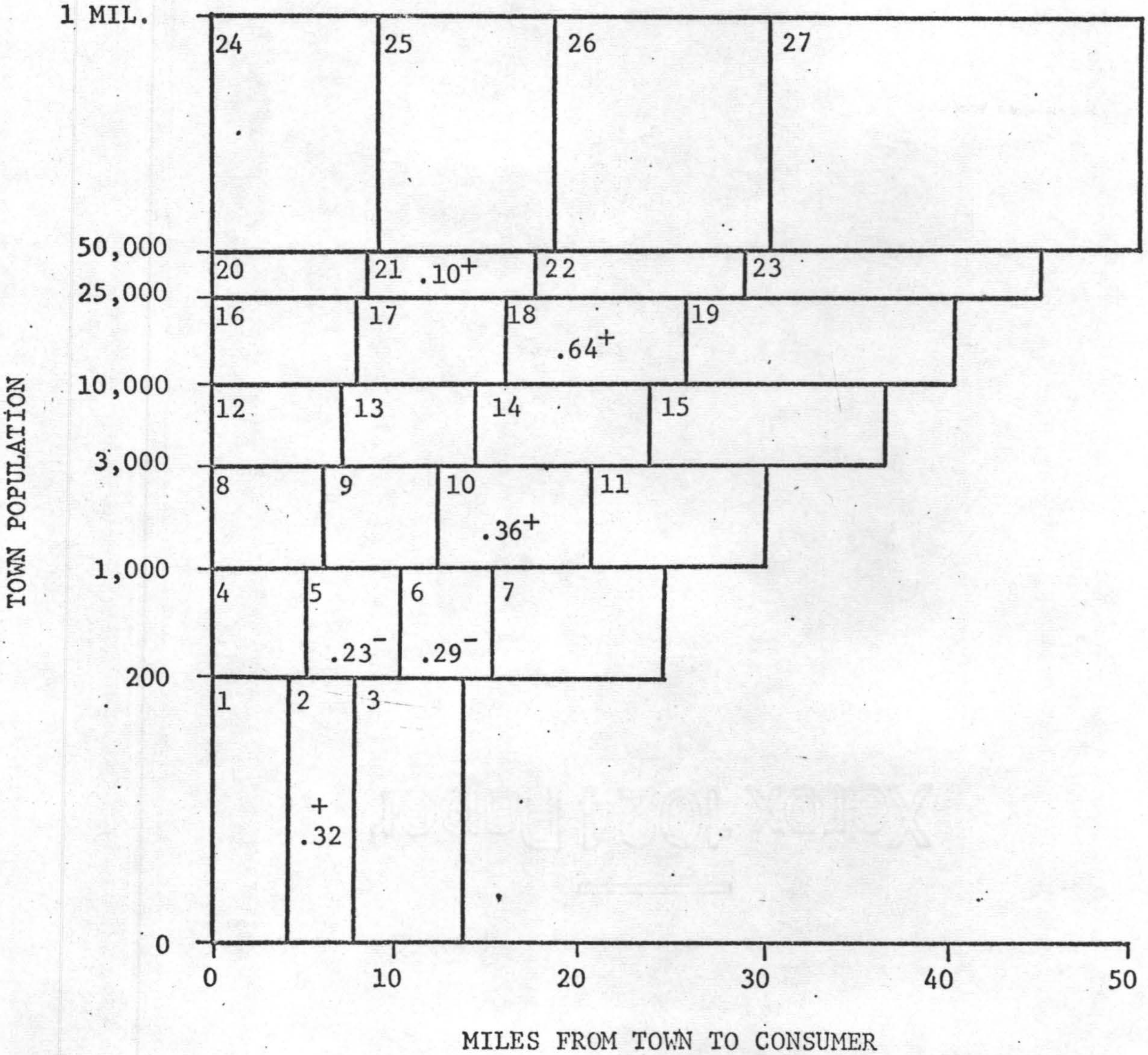
Four Largest Significant Difference Proportions for Each Location Type in 50 Random Groupings

Location type	Proportions			
1	.18	.18	.24	.59
2	.27	.32	.32	.36
3	.14	.14	.14	.14
4	.18	.18	.18	.27
5	.18	.23	.32	.38
6	.23	.23	.27	.32
7	.10	.10	.14	.15
8	.05	.05	.05	.10
9	.18	.18	.18	.23
10	.32	.36	.41	.48
11	.14	.14	.18	.18
12	.00	.00	.00	.00
13	.14	.14	.14	.14
14	.24	.27	.36	.50
15	.27	.27	.36	.45
16	.00	.00	.00	.00
17	.05	.09	.09	.15
18	.18	.32	.32	.41
19	.00	.00	.00	.00
20	.00	.00	.00	.00
21	.05	.05	.10	.21
22	.14	.27	.27	.32
23	.15	.15	.16	.17
24	.00	.00	.00	.05
25	.05	.05	.05	.05
26	.14	.18	.27	.27
27	.14	.18	.23	.53

Figure 5.4

Location Types "Preferred" Significantly More or Less

by Households with One or More Members Working Off the Farm



\+ indicates "more preferred" to a significant proportion of other location types.

\- indicates "less preferred" to a significant proportion of other location types.

Proportions indicate the proportion of times the location type is significantly differently preferred.

three and never the last one. Therefore no subsets of the sample can differ in preference for these types. In the remaining location types the absence of significant differences in any great number may truly reflect the absence of preference differences for these location types in the population which each group represents. Alternatively, in some of the statistical tests the null hypothesis may have been wrongly accepted, whereas a larger sample would reduce the chance of this and show other significant preference differences.

The general pattern and direction of significant preference differences suggests households with members working off-farm "prefer" more distant, larger places and smaller, nearby places, whilst other households have a greater preference for places intermediate to these other types. However, by one interpretation of this difference the phrase "spatial preference" is inappropriate in this situation. Those householders whose daily work takes them off the farm, very probably work in or pass through towns. Since larger places are larger employment sources and since people are generally willing to commute further to earn money than to spend it, it is probable that many of them work in larger and more distant places than they might otherwise shop in. As a result, it is feasible that some grocery and other purchasing is done by the off-farm worker at or on the way to the work place. A complementary response to the household's opportunity to shop in larger places is that the household needs only patronise very local places for occasional needs. By contrast, the household which lacks the advantage of someone regularly visiting a "large" place is more likely to patronise a smaller, less distant place and as a result have less need to patronise a very local place.

The above analysis of 12 pairs of socio-economic groups is repeated using different watershed values since it is arguable that preference variation may be a discontinuous function of socio-economic attributes, and therefore more discernible with some watershed values than with others. The range of possible watershed values for each attribute is limited by the problem of one group being too small for effective statistical comparison of preference proportions. The watershed values chosen are shown in Table 5.7. Table 5.8 shows the percentages of

Table 5.7

Other Watershed Values Used to Group Households by Socio-Economic Characteristics

Socio-economic attribute		1	2	3	4	5	6	7	8	9	10	11	12
Watershed values	1	2	0	0	36	8	37	8	0	100	10	8	1,500
	2	3	1	1	43	10	44	10	1	175	16	13	2.300
	3	4	2	2	50	12	51	12	2	250	22	18	3,100

Table 5.8

Percentages of Significantly Different Paired Comparisons in Socio-Economic Groups' Preference Structures, sing Three Different Watershed Values for Each Variable

Groups defined in terms of variable no.		1	2	3	4	5	6	7	8	9	10	11	12
% of significant differences using each of the three watershed values	1	5	9	9	8	5	7	6	14	4	4	7	3
	2	6	7	5	6	8	10	4	6	4	9	10	7
	3	8	6	8	10	5	8	5	16	1	7	4	4

significantly different paired comparisons achieved with each of three watershed values for each variable. No other pair of groups is revealed to have significantly different preferences. The explanation of the percentage variation for variable 8 is that there are few households with more than one member working off the farm and very few with more than two. Hence the percentages on the second and third iterations are based on the comparison of very few p estimates, resulting in the large variance in significant differences. Therefore, within the limits set by sample size, no significant increase in preference differences is obtained using other watershed values.

It is arguable, however, that preference differences are more likely between groups whose socio-economic characteristics are very different. For example, very young and very old households may have more marked spatial preference differences than simply younger and older households, whose mean age difference may only be 15 years. In other words, the relationship between preference and socio-economic characteristics may be observable in a sample only when extremes of socio-economic types are compared. This hypothesis is tested by comparing the preferences of only those households who have relatively extreme scores on a socio-economic variable. The upper score for the lower group and lower score for the upper group for each of the 12 variables is shown in Table 5.9.

None of the groups has significant preference differences which could not be attributed to chance. In the case of the one pair of groups already found to have significant preference differences, the most likely explanation of an absence of significant difference in this case is that one group contains only 66 households.

Table 5.9

Upper and Lower Variable Scores of "Extreme" Socio-Economic Groups and the Sample Size of Each Group

Variable	Upper limit	Lower limit	Sample size	
			Lower group	Upper group
1	2	5	117	160
2	0	3	231	75
3	0	3	240	39
4	35	49	132	154
5	8	11	124	277
6	36	50	115	181
7	8	11	192	197
8	0	2	261	66
9	100	250	86	126
10	10	20	150	184
11	9	18	173	171
12	1,500	3,000	109	157

Table 5.10

The Highest Loading Variables on Each of 5 Factors

Factor	Highest loading variables	Factor loadings
1. Youthfulness of household and residency.	Years operating this farm	-.846
	Age of farm operator	-.835
	Age of homemaker	-.796
	Years living in this house	-.782
2. Adolescent/ off-farm workers	Number of householders working off farm	.858
	Number of persons between 10 and 20	.636
3. Prosperity	Number of acres	.901
4. Education	Homemaker's years of education	.848
	Farm operator's years of education	.771
5. Family size	Number of persons in household	.923
	Number of persons between 0 and 10	.835

Preference Comparisons of Groups Defined by Households' Factor Scores

A further hypothesis tested is that households are more adequately characterised by attributes common to more than one variable. In other words, it is hypothesized that households are characterised by more basic dimensions than the variables discussed, and that households' scores on these more fundamental attributes are more likely to explain preference differences. In order to derive these attributes from the existing variable scores on 11 of the 12 variables,[6] a principal component factor analysis with orthogonal rotation is used. The 11 original variables most appropriately collapse to 5 dimensions, which are labelled in Table 5.10. On each factor, one to four variables load very highly (over 0.77 in all cases but one) and the remainder have low loadings exceeding 0.38 in only one instance. As with variable scores, so are households' factor scores on each factor used to group households. For each factor in turn, households are assigned to one of two groups depending on whether the factor score is greater than or less than 0. None of the five percentages of significant differences are higher than might be attributable to chance. As in the test on variable scores, the watershed is varied to -0.5 and +0.5 on each factor, but in no case is the percentage of significant differences between two groups higher than 10%. A comparison of groups characterised by extreme factor scores (less than -0.5 and greater than 0.5) results in no increase in the percentage. Thus the hypothesis is rejected that a factor characterisation of households reveals a significant amount of preference differences.

---

6 Missing information is too frequent on the 12th variable, net farm income, to include it in the factor analysis.

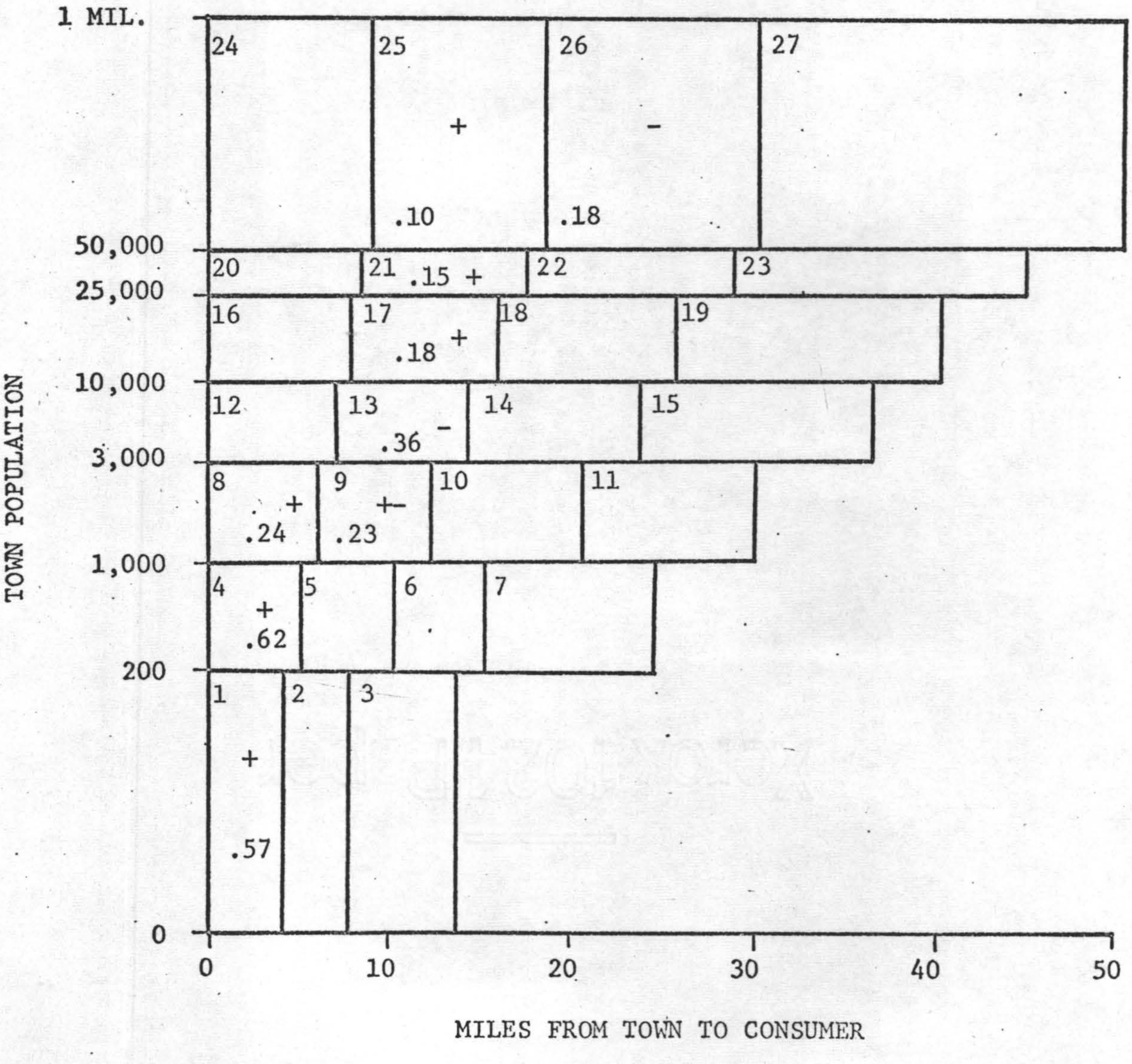
Preference Comparison of Groups Making Different Spatial Choices

In Chapter 2 reference is made to Michelson's suggestion (1966) that general values and preferences are reflected in people's spatial preferences. For example, he finds that people who seek class exclusiveness also desire spatial exclusiveness for a class. Thus, differences in spatial preferences may be viewed as revealing value differences. This notion is tested here in a limited fashion, since few hypotheses can be made about people's values from the limited data available. However, it is hypothesized that households which value time more highly will prefer convenient places more than other households. The operational hypothesis tested is that households whose maximum grocery purchase town is the nearest have different preferences from other households. Even if the latter hypothesis is accepted, it can only be surmised that this preference for convenient places is attributable to the higher value of time. The 33% of households maximally patronising the nearest town do indeed have significantly different preferences, with a 17% significant difference between the two groups. Those location types whose proportion of significant differences exceeds the fourth largest proportion for random groupings (see Table 5.6) are shown in Figure 5.5. Indeed, it is noteworthy that of the 9 location types satisfying that condition, 5 exceed even the largest random grouping proportion for that location type and 3 more exceed all but the largest proportion. Therefore, the preference difference between the two groups is much stronger and more widespread in terms of location types affected than the significant differences related to off-farm workers.

The pattern very clearly indicates that the difference between house-

Figure 5.5

Location Types Preferred Significantly More or Less by Households Maximally Patronising the Nearest Town



+ indicates "more preferred"; - indicates "less preferred"; +- indicates "more preferred relative to some location types, less preferred relative to others".

Proportions indicate the proportion of times the location type is significantly differently preferred.

holds patronising the nearest place and those not, is to be explained by more than just the opportunistic location of households in the former group. The analysis shows that even when similarly located with respect to spatial alternatives, certain households have a greater preference for nearby locations. Inevitably this difference in "nearness" preference manifests itself only with regard to smaller towns, since for large nearby towns, any difference in the weighting on distance is obscured by the effect of a large population on the household's preference calculation. However, the nearness preference is revealed for larger places in the second nearest distance categories for all towns larger than 10,000, i.e., for location types 17, 21 and 25. As a result of this emphasis on nearby places, it is significant that amongst towns less than 10,000 the slightly more distant location type 13 is less preferred by the group patronising nearest towns. This result would appear to agree with the intuitive notion that a location type which is just a bit larger but also a bit more distant than types 1, 4, 8 and part of 9 is most likely to suffer as a result of their gains. Thus, the loser in the spatial competition is location type 13. Likewise, amongst towns over 10,000 location type 26 suffers vis-à-vis the gains of types 17, 21 and 25 in exactly the same fashion as 13 suffers vis-à-vis the gains of 1, 4, 8 and 9.

By contrast, location type 9 is both loser and winner. With regard to location types 1, 2, 4, 5 and 8 it shares the fate of type 13. But 9 is relatively more preferred to 13, 21, 22, 25, 26 and 27, though not always significantly, by those preferring nearby places than by others.[7] Thus, it lies astride the boundary separating location types more and less preferred by households preferring nearby places. Where nearer places are available, 9 loses trade, but where it is nearer than other larger places, the group's "nearness" preference operates in 9's favour.

The exceptions to this "nearer types are relatively more preferred, further are less preferred than 9" generalisation are location types 14, 17 and 18 to which 9 is less preferred by the group preferring nearer places. The explanation may be that location types 14, 18 and 19 complement the more preferred small, nearby places. In other words, if the maximum grocery purchase town belongs to location types 1, 4 or 8, less frequent purchases may be made at a larger, more distant place such as types 14, 17 and 18, which supply goods not commonly obtained in the smaller places. Thus, the fact that location types 14, 17 and 18 rank higher vis-à-vis 9 on the nearness-preferring group's scale may simply reflect their complementary role to small, nearby places.

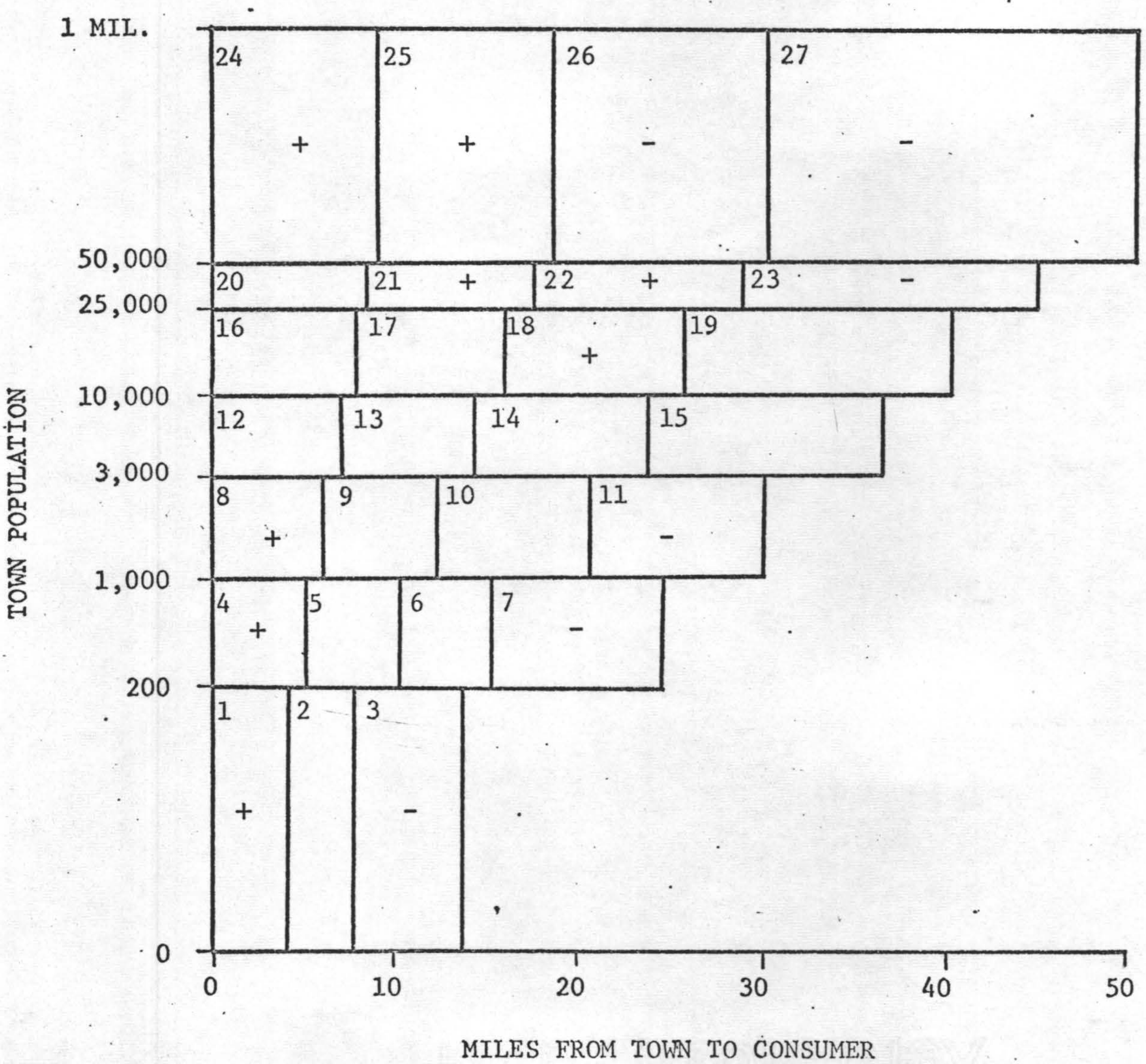
The overall preference of one group for nearer places is illustrated more vividly in Figure 5.6. The general direction of preference differences irrespective of statistical significance is obtained from a matrix of differences in preference proportions. Location types are indicated as more or less preferred by one group if more than 75% of all preference differences for a particular location type are in the same direction.

---

7 It should be noted that a location type which is relatively more preferred to certain location types by one group than another simply lies higher on one group's preference scale. It does not necessarily mean the location type lies above these others on the one group's scale.

Figure 5.6

Major Direction of Preference Differences for Each Location Type with More Than 75% of Preference Differences in the Same Direction



Particularly regarding towns less than 10,000, one group's relatively greater preference for near places and lesser preference for distant places is clear. For larger towns, the nearest towns are equally highly preferred by both groups excepting location type 24. Thus for towns greater than 10,000, preference differences are more notable only at intermediate and far distances.

Best-fitting gravity models of the kind described by equation 4.21 yield the following results for the two groups - $\frac{p^{.32}}{d^{2.2}}$ for the group preferring smaller, nearby places, and $\frac{p^{.32}}{d^{1.0}}$ for the other group. The greater disutility attached to distance by the former group is to be expected. However, the difference between the functions may result from each being derived for a different number of location types. The different number is a consequence of the previously discussed problem of the scaling algorithm being unable to scale location types whose proximity to all other types is invariant, in this case 0.5. Obtaining utility scores for only those location types scalable for both groups, results in the following gravity models - $\frac{p^{.32}}{d^{2.2}}$ for the first group and $\frac{p^{.53}}{d^{2.2}}$ for the other. Only 13 location types are involved. All but three types above the 10,000 population limit (18, 26 and 27) are omitted, while below that population, some of the nearest types (1 and 12) and of the farthest (3, 7 and 11) are omitted. Within these limits, the two groups attach the same disutility to distance, but the group preferring smaller, nearby places reveals its lower preference for larger places by its smaller p exponent.

Having defined two groups with differing space preferences, a statistical test on the variable scores in each group is performed. The null hypothesis tested for each of the 12 variables is that the sample means for each group can be regarded as samples of the same population mean. The appropriate rejection region for the statistic t is $-1.64 > t > 1.64$ at the $\alpha = 0.1$ level. In all three cases the null hypothesis is rejected at the $\alpha = 0.1$ level (see Table 5.11).

Table 5.11

Socio-economic variable	t value	Sample mean for group preferring nearby places	Sample mean for other group
Number of persons in household	-2.046	3.8	4.2
Number of persons from 0 to 10 years old	-1.718	.95	1.2
Net farm income in dollars	1.695	2875	2470

The conclusion is that smaller, more affluent families with fewer young children prefer to patronise more nearby places than other families. This could be taken to indicate that the small affluent family emphasizes convenience, whereas the larger family with young children and a smaller budget, emphasizes savings through comparison shopping. Comparison shopping could commonly involve trips further afield. It is also noteworthy that apart from "the number of children 10 years old or less", the most highly correlated variables with "family size" are "age of homemaker" ($r = -0.45$) and "age of farm operator" ($r = -0.39$). Thus, although the two groups are not characterised by a statistically significant difference in average ages of homemakers and farm operators, it seems likely from the above correlations that a higher than average preference for nearby places may also be characteristic of older households. Thus, for example, the family in which the children have grown up and left home, being relieved of the financial burden of children, might well prefer convenience to savings. Put another way, families with younger children are normally at that stage in the life cycle where non-convenience good purchasing is at a maximum. Thus, such families are more likely to engage in grocery purchase trips in which groceries are not the highest order good sought. Therefore, they are more likely to patronise larger places which are usually not the nearest town. By contrast, older families, especially those whose children no longer live at home, demand high order goods less often, so that the goods they seek on a grocery trip may be provided more often by the nearest place.

It should be noted that in the initial groupings on the basis of variable scores, no significant preference differences are noted for the three variables subsequently shown to be related to preference variation. It would appear there are sufficient smaller, more affluent families without a higher than average preference for nearby places, for no significant preference difference to show up when households are grouped by socio-economic attributes. However, when grouped by town patronage the socio-economic differences in the two preference groups are revealed. It is therefore not necessarily valid to conclude that a socio-economic variable is unrelated to preference variation solely on the basis of comparisons of socio-economic groups' preferences, contradictory as the statement may seem. Also, it appears from all of the foregoing group preference comparisons that grouping by spatial behavioral traits is a more powerful method for revealing spatial preference variation than grouping by socio-economic attributes.

Preference Comparison of Multivariately Defined Groups

Despite the above conclusion, one further hypothesis involving socio-economic grouping is tested. It is hypothesized that a multivariate grouping of households in terms of the three variables shown to be related to preference variation, may result in significant preference differences between the groups. Specifically,one group is characterised by scores less than the watershed values shown in Table 5.1 on all three variables 1, 2 and 12, the other by scores greater than all three watershed values. Although preference differences between the two groups are not significant it is impossible to say whether or not this is the result of each group having fewer than 70 members. In cases of very small samples, preference differences are very unlikely to be revealed, since so many matrix cells lack data.

Comparison of Preferences Based on First and Second Town Choices

In line with the earlier hypothesis that differing goals may be reflected by differing space preferences, two further hypotheses are tested. It is hypothesised that first choice towns and second choice towns, as defined by dollar expenditure, satisfy different goals and therefore that the location types preferred as maximum purchase towns will differ from those preferred as second choice towns. Essentially, this is a test for complementarity in spatial preferences. Two paired comparison matrices are derived, one based only on paired comparisons related to households' first choice location types and the other based on comparisons related to second choice location types. However, only 10% of all paired comparisons are significantly different. Whilst such a percentage might be attributable to chance, it is noteworthy that two location types are significantly differently preferred in a proportion of cases greater than in any of the random groupings, namely type 7 (0.33) and type 18 (0.52). The direction of the differences is consistent and indicates 7 is more preferred as a first choice town and 18 more preferred as a second choice town. This indicates, not surprisingly, that a small distant town whilst ranking low both as a first and second town choice, is even less likely to be chosen as a supplementary grocery supply centre than as a primary supplier. This is because the role of secondary supply centre is more than likely taken by more distant large places in which grocery shopping is a secondary goal. Indeed, this would seem borne out by 18's greater preferredness as a second choice place.

The absence of an overall significant amount of preference difference between the two groups need not be regarded as a rejection of the hypothesis that different location types are preferred as first and second choice towns, at least at the individual level. The aggregation of households may cancel out complementaries in location type preferences for first and second choice places which may only be observable at the level of the individual household. Thus, some households may prefer small, nearby places as first choice towns and complement this with a large, more distant second choice town. Likewise, other households may have an opposite complementary preference. But if two such sets of households are combined in one sample, the one set of preference complementaries cancels out the other, so that no great difference between first and second town preferences is revealed.

Preference Comparison of Households Patronising One and More Than One Town

The final hypothesis tested is that households who patronise only one town for groceries have different space preferences from households patronising more than one town. The hypothesis is rejected with only 5% significant differences between the two groups' paired comparison proportions.

Test of the Method's Sensitivity in Discriminating between Groups' Preferences

To put the above results in perspective, tests are made to determine how small preference differences can be distinguished by the method used above. Using the hypothetical data sets described in Chapter 4, each sample is divided into two groups according to its known preference rules.[8] The proportion of significant differences between each pair of groups and the proportion of times each location type is significantly differently preferred, are shown in Table 5.12.

Of the first four groupings the method can distinguish significant amounts of difference between groups with the two most different rules, but not between groups with the two least different rules. It is notable that neither the deterministic data set (3) nor the probabilistic set (10) with the same preference rules contains significant differences. Consistent with its inability to distinguish between the 3rd and 4th

---

8 Where three preference rules exist, one group contains households obeying the first two rules as defined in Table 4.5, and the other group contains households obeying the third rule. Where rules are normally distributed one group contains households obeying rules less than or equal to the mean rule, the other group contains the remainder.

Table 5.12

The Proportion of Statistically Significant Preference Differences for each Location Type by 11 Pairs of Groups with Known Hypothetical Preference Rules

Location types	Pairs of groups 1	2	3	4	5	6	7	8	9	10	11
1	.52	.38	.14	.10	.15	.14	0.00	.30	.10	.25	.35
2	.84	.33	.18	0.00	.05	.05	0.00	.11	.05	.14	.23
3	.56	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	.05	.09
4	.59	.18	.14	.14	.14	.18	0.00	.50	.29	.23	.36
5	.71	.36	.27	.18	.14	.23	.23	.36	.23	.32	.36
6	.36	.10	0.00	0.00	.05	0.00	0.00	.11	0.00	0.00	.14
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	.09
8	.30	.05	.05	.05	.05	0.00	0.00	.19	.10	0.00	.05
9	.52	.32	.14	.05	.09	.14	.11	.18	.18	.14	.23
10	.56	.14	.05	0.00	0.00	.05	0.00	.23	.05	.10	.23
11	.33	0.00	0.00	0.00	0.00	0.00	0.00	.10	0.00	.14	.27
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	.44	.20	.14	.10	.09	.09	.05	.23	.10	.05	.09
14	.78	.21	.10	.05	.05	.09	.05	.24	.14	.09	.23
15	.78	.12	.11	.06	.09	.06	0.00	.32	.23	.27	.45
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17	.50	0.00	0.00	0.00	0.00	0.00	0.00	.05	0.00	.05	.05
18	.43	.19	.05	.05	.05	.16	.06	.10	0.00	0.00	.14
19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21	.50	0.00	0.00	0.00	0.00	0.00	0.00	.13	0.00	.12	0.00
22	1.00	.24	.22	.11	.06	.24	0.00	.10	.05	.11	.10
23	.89	.36	.11	.05	.10	0.00	.17	.30	.15	.11	.35
24	.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25	.80	.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
26	.88	.20	.11	0.00	.05	0.00	.06	.15	0.00	0.00	.10
27	.88	.67	.26	.16	.20	.17	.06	.35	.25	.16	.20
	.59	.18	.09	.05	.06	.07	.04	.19	.09	.10	.19

The final row shows the total proportion of significant differences between each pair of groups. The pairs represented in each column from left to right are:

1. d min. (.5); p/d max.(.5)
2. p/d^4 max. (.5); p/d^2 max. (.5)
3. p/d^3 max. (.5); p/d^2 max. (.5)
4. $p/d^{2.5}$ max. (.5); p/d^2 max. (.5)
5. p/d^3 max. (.33); $p/d^{2.5}$ max. (.33); p/d^2 max. (.33)
6. p/d^3 max. (.2); $p/d^{2.5}$ max. (.2); p/d^2 max. (.6)
7. p/d^3 max. (.1); $p/d^{2.5}$ max. (.1); p/d^2 max. (.8)
8. p/d^2 max. normally distributed with d's exponent variance = 1.
9. As in 8, but with d's exponent variance = .5.
10. As in 3, but probabilistic.
11. As in 8, but probabilistic.

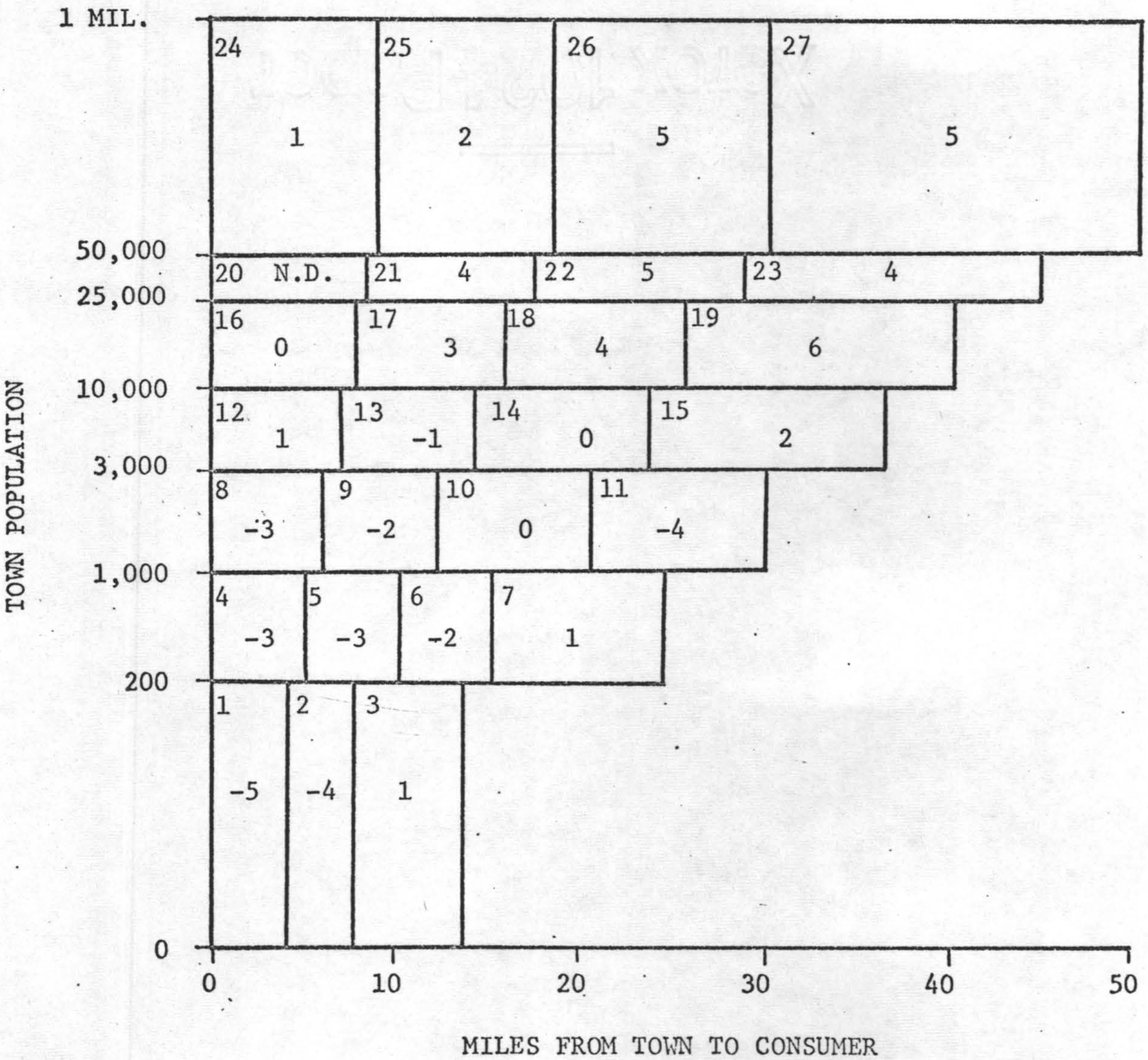
pairs of groups, the method does not show up marked difference between the 5th, 6th and 7th pairs, none of which have more extreme rules than the third pair's p/d^3 and p/d^2. In the 8th data set, and in the 11th, where d's exponent variance is 1, significant differences are observed, where one set is probabilistic (11) and the other deterministic (8). However, a reduction of the variance to .5 in the 9th set results in only 9% significant difference compared to 19% with a variance of 1 in the 8th set.

It can be concluded that fairly similar preference rules such as p/d^3 and p/d^2 and even more similar rules cannot be distinguished using this method with the given significance criteria and <u>with this sample size</u>. Thus, those groups shown to be different in the real sample analysis have more than just a small difference in their preference structures.

A simple comparison of the ranking of the two preference groups in sets 3 and 4 does, however, distinguish between the two groups in each set, sufficiently to be able to discern that the group in each set with the higher d exponent ranks larger and more distant places lower than the other group, and consequently ranks small nearby places higher (see Figures 5.7 and 5.8). Thus, the effect of the higher d exponent is revealed in those groups' lower preference for more distant places. The conclusion to be drawn is that, although the rigorous demands of the statistical significance criteria do not necessarily show up the smaller preference differences with this sample size, the differences can be distinguished using the statistically less reliable rank differences method. Thus, the method of paired comparisons itself is

Figure 5.7

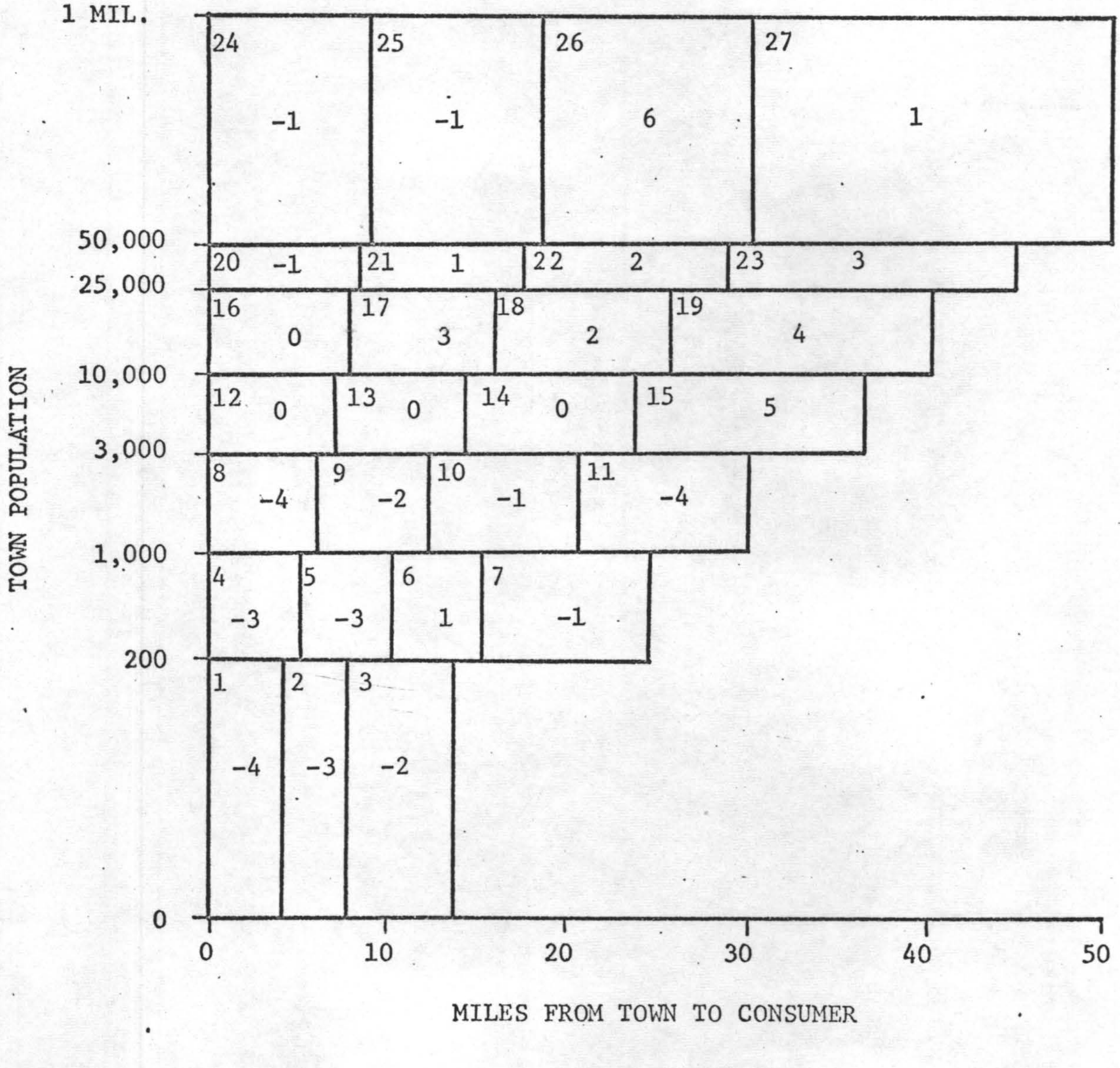
Location Types' Rank Differences (p/d^3 Rank - p/d^2 Rank)



N.D. signifies no data available for one or both groups.

Figure 5.8

Location Types' Rank Differences ($p/d^{2.5}$ Rank - p/d^2 Rank)



highly sensitive even to small preference differences. Since the members belonging to each preference group in the hypothetical data sets are known, no risk is run using this method. However, where the preference rule of each member of a real sample is not known, the risk with the rank difference method, as was discussed earlier in this chapter, is that of drawing spurious conclusions about preference differences between randomly different groups, since no statistical criterion is available to indicate what amount of rank difference can be attributed to chance.

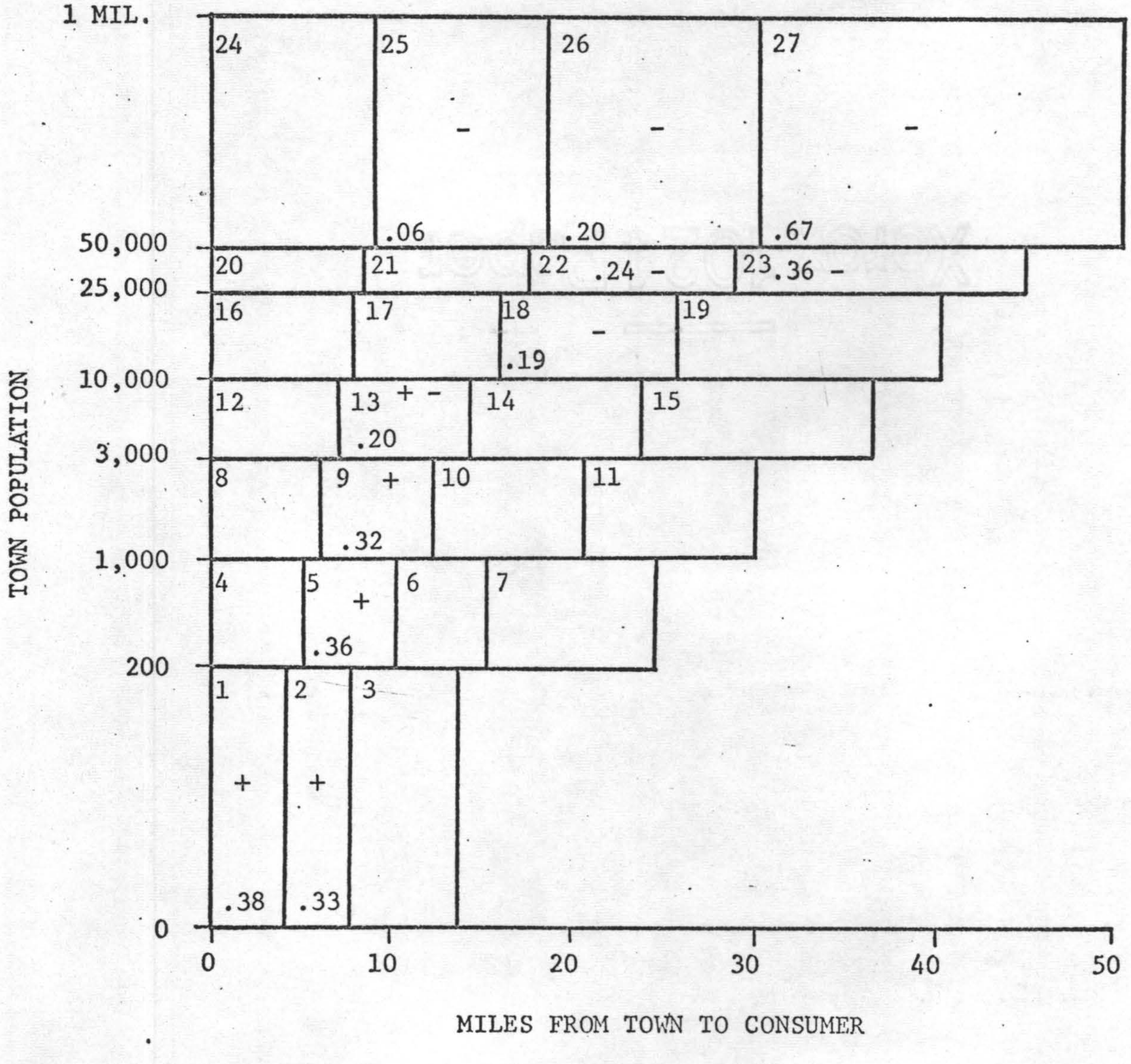
As well as being able to distinguish the preference difference between two groups, the results of the statistical tests, which are summarised in Table 5.12, enable the nature of the preference difference to be determined. Figure 5.9 shows the direction of preference differences for location types significantly differently preferred by the groups maximising p/d^4 and p/d^2. From the figure it is clear that the former group prefers nearer places and therefore can be inferred to assign greater weight to distance than the other group, as is in fact the case. As is shown in Chapter 4, the information in the preference matrix for each such group is adequate to determine the relative weightings assigned by each group to p and d. Therefore, the method of paired comparisons together with Guttman-Lingoes SSA-I enables a relatively accurate reconstruction of the initial preference rule of such groups. Any inaccuracy would seem from the present analysis, to be more a function of sampling error and the aggregating effect of location types than of the above two methods.

Summary

In summary, several points are worth stressing. Firstly, regarding

Figure 5.9

Location Types Preferred Significantly More or Less by Households Maximising p/d^4 Than by Those Maximising p/d^2



+ indicates "more preferred".

- indicates "less preferred".

Proportions indicate the proportion of times the location type is significantly differently preferred.

the socio-economic variables considered, it would appear that the correlation between socio-economic scores and preference rules is not sufficiently strong for groupings based on socio-economic scores to show up preference differences associated with these characteristics. By contrast, groupings based on households' spatial behavior have proven more powerful in showing up preference differences, which were subsequently linked to certain socio-economic household attributes. One tentative conclusion is that a characterisation of households by spatial behavior would seem a more fruitful approach to grouping, for the purpose of discerning preference differences. Therefore, in addition to collecting socio-economic data on consumers in future, it would seem worthwhile to collect more data on their spatial behavior. Thus, for example, information on trip frequencies to different centres and the mix of goods and services purchased at each, would supply useful data for characterising households behaviorally.

Regarding the method used to compare groups' preferences, it seems the method of paired comparisons used in conjunction with the statistical tests described is sufficiently sensitive to show differences for all but small preference differences such as p/d^3 and p/d^2. However, since there is no way to know what magnitude of preference difference is significant either in terms of behavior or of central place analysis, it is impossible as yet to judge whether this level of refinement is adequate. Therefore, it would be worthwhile either to seek larger data sets or, lacking this, to pursue statistically less rigorous techniques for preference comparisons, such as the method of rank differences, if a suitable test of statistical significance could be devised.

CONCLUSION

The purpose of this study has been two-fold. The first was to determine the amount and nature of lawfulness in consumers' spatial preference functions, with regard to towns for grocery purchases. The second was to determine if space preference functions differ significantly between households in any orderly fashion. A review of geographical and other literature revealed serious weaknesses in most of the few studies aimed at one or other of these goals. In the one set of studies by Rushton (1969a, 1969b) which come closer than any others to satisfying the above goals, their scope is more limited than this study's, and the reliability of some of the analytical methods was questioned.

In the most general terms, the conclusion reached here is that there is considerable lawfulness in consumer space preferences, but that different preference rules characterise different types of consumer. Whilst a unique assignment of preference rules to household types was not sought, it is now at least possible to say that certain household types have a higher probability of conforming to one preference rule than another.

More specifically, in Chapter 4 it is shown that no major explanatory variable has been omitted from the model, given the high transitivity computed using the present model. Thus, it seems reasonable to conclude that a general model of central place preference should not require any more complex a definition of spatial alternatives than that used here.

Although the aggregate preference proportions are concluded to have ratio properties, analysis of hypothetical data shows these properties to be partly a function of the similarity in the preference rules of members of a data set, irrespective of whether the individuals have probabilistic or deterministic rules. In addition, proof of a ratio scale seems to be a function of the scale of analysis, in particular the assignment of many population/distance combinations to the same location type. If so, any future test for individual cardinal utility functions should avoid the present study's kind of operational definition of spatial alternatives. No empirical proof, therefore, is provided as to whether individuals have deterministic or probabilistic preference functions.

Despite the disadvantage of the scale of analysis, it does appear to have an accompanying advantage in the event that preference is deterministic. The scale effect on preference probabilities is that values other than 1, 0 and .5 are possible, whereas, if preference is truly deterministic, and the scale effect removed, only the above three values, signifying absolute preference and indifference, would be possible. In the latter case a cardinal preference scale, such as that derived using Guttman-Lingoes SSA-I, could not be obtained with such values and instead only an ordinal preference function could be derived. As a result it would be impossible to compute the weightings assigned to population and distance in a preference function. The latter computation is of course possible, given more than the above three probabilities in a preference matrix, which, as the hypothetical data sets show, is possible with deterministic preference functions given the present scale effect in defining spatial alternatives.

Proof of a ratio scale is also shown to eliminate the necessity of collecting information on pairwise preferences for every pair of alternatives. Instead, it is possible to compute all n(n - 1)/2 paired comparison proportions from a minimum of n - 1 proportions.

In addition, it is shown that a ratio preference scale has more useful predictive properties for central place analysis than a deterministic ordinal preference scale. The latter assigns all expenditures from a point to the same most preferred alternative available. By contrast, the probabilistic properties of a ratio scale enables different proportions of expenditure from a point to be allocated to competing alternatives. This less extreme allocation of trade to centres would seem more realistic and hence more useful in empirical central place analysis.

Regarding the likely range of preference rules, discussed in Chapter 4, the answer depends on whether individual preference is deterministic or probabilistic. If the former, the evidence is that the preference rules have a fairly limited range. But if probabilistic, it is impossible to estimate the range of rules present in the sample, since the possibility of so many different types of probability function limits the inference to be made by comparing the real data set with hypothetical samples based on only one type of probability function.

In Chapter 5, it is shown how proof of ratio properties in the preference matrix can be used to obtain improved p estimates.

A comparison of socially and economically defined groups results in little difference being observed in their space preferences, except in the case of households with members working off the farm. In that case, these families "prefer" or more likely have a better opportunity by virtue of work-related travel, to patronise larger, more distant places than other households. In contrast to the relative absence of differences when households are grouped socially and economically, marked preference differences are revealed when households are grouped behaviorally according to whether they patronise the nearest place or not. Instead of discovering, as might be hypothesized, that those patronising the nearest place are simply more opportunistically located by having a larger than average alternative nearby, it is shown that the group does in fact prefer nearby places more than the other group. This former group is shown to be characterised by significantly higher incomes, smaller households and fewer young children than the other group. The preference difference reflects perhaps the different goals of the two groups, given that the less affluent households with more young children and a larger family would likely be more budget conscious, and therefore more likely to engage in comparison shopping, which commonly involves trips to more distant places. In addition, young children in a family tend to coincide with the time when more non-convenience goods are sought, so that larger, usually more distant, towns are more likely to be patronised.

A general, though somewhat tenuous conclusion, based on the results in Chapter 5 is that clues as to preference differences would seem more likely to be gained from observing overt behavior rather than social and economic variables. This coincides with Michelson's finding (Michelson, 1966) that socio-economic variables were poor explainers of preference variation but that more subtle variables such as more general values, as well as behavior, seemed better indicators of spatial preference variation.

Finally, the method of paired comparison combined with statistical tests is fairly sensitive to preference differences in hypothetical data. Any inadequacy in preference discrimination seems to be more a function of the scale of the analysis and the sample size than of any innate deficiency in the method. Thus, the continued use of this method for comparing preference structures seems justified.

Future Research Problems

Probably the most pervasive stumbling block in this analysis is the generalisation of spatial alternatives into 27 location types. Future analyses would be better designed without this aggregation. This should be possible in the future, certainly if ratio scale preference can be proved, and probably even with only knowledge of the general order of preference. Using the ratio case as an example, if only the approximate shape of a preference function is known, as it is from this study, n discrete population/distance combinations can be selected, p/d combinations for short. Knowing their relative positions on the approximate ratio preference scale, or their order on an ordinal scale, a population can be sampled such that only information on the proportionate preference between the n - 1 pairs of adjacent p/d combinations on the ratio scale need be sought. From these proportionate preferences, the distance between adjacent points on the ratio scale can be calculated. Having defined the ratio scale, the distance between and therefore the amount of preference between all pairs of points can be readily computed. Therefore, only a very few pairs of specific p/d combinations would be needed to accurately derive the preference function. It should be possible to find appropriate pairs of specific p/d combinations which exist in reality and to intensively sample residents so located. This can be repeated for as many pairs of p/d combinations as thought desirable, since the more p/d combinations that are assigned preference scores, the more accurate a preference function can be fitted to the set of scores for p/d combinations. Thus, it would appear possible in future analyses to remove the current generalisation of spatial opportunities and the associated inaccuracies. Since the data used in this study were not collected with the above objectives in mind, few, if any, households compare the same pair of p/d combinations. Thus, it is necessary to aggregate many p/d combinations into the same location type, in order to derive an approximate preference function.

In light of this study's inconclusive findings regarding ratio-scale preference, an analysis properly designed to test for this at the level of individuals would seem an obvious progression. The knowledge gained would have both intrinsic value in terms of human spatial preference, as well as being useful in formulating more realistic preference premises in central place analysis. Just as in the previous discussion, current random samples would appear inappropriate for such a test. Once again, intensive samples in specific locations which meet the conditions of a test for ratio-scale preference would be required.

Other more obvious and simple extensions of the present work are apparent. The analysis might be repeated for other commodity purchases and the nature of difference in preference structures be observed. Given a space preference function for each good and service, it would be possible to compute for each town in any given area, such as Iowa, what functions the town could economically support, in light of the threshold requirement of each. The result could be compared with reality. Presumably such an exercise would be useful in determining the possible future pattern of central place growth and decline in an area, if we accept the assumption that the pattern of supply centers is a response to consumer spatial preference as revealed in spatial demand.

Finally, although individual preference has not been shown to have ratio-scale properties, these have been shown to characterise the aggregate preference structure. The problems of incorporating such a preference function into a central place model, to replace Christaller's distance minimising behavioral postulate, and determination of the spatial consequences of such a change would seem an obvious application of present results in the effort to make the premises and predictions of Central Place Theory more realistic.

APPENDIX A

Description of the Sample

INTRODUCTION

In the spring of 1961, a survey of households and farm expenditures and sales by persons living in rural Iowa was conducted by the Iowa State University Statistical Laboratory for the Iowa College-Community Research Center.[1] The purpose of the study was to measure the economic impact of the expenditure patterns of these people on towns of various sizes and at various distances and to gain some insight into the probable effects of continued decrease in the rural population of the state on these types of communities.

THE UNIVERSE

Two units of observation, households[2] and farms,[2] were recognized in this study. The universe sampled included all households located in the open country zone of Iowa and all farms operated by persons living in these households. The open country zone, as defined by the Master Sample of Agriculture, consists of all land area outside the boundaries of incorporated towns and cities, unincorporated name places, and built-up areas near cities having a population density of 100 or more persons per square mile. The boundaries and, in the latter case, the population density, are defined as of 1940.

THE SAMPLE DESIGN

In order to make the territorial distribution of the farms in the sample as broadly representative as possible, the sample was allocated to the counties in Iowa in proportion to the total number of farms in each county according to the 1959 Census of Agriculture. Although the sample size per county (4 to 13 segments) is too small for making individual county estimates, estimates for aggregations of counties (e.g., types of farming areas, census economic areas) can be made.

Because it was assumed that closely grouped farms would tend to have

---

* This appendix, a preliminary description of the sample methods, was written by Professor Strand and his staff at the Statistical Service Division, Iowa State University, Ames.

1 The Iowa College-Community Research Center is composed of Iowa businessmen and selected research personnel of Iowa State University and The University of Iowa.

2 These terms are defined elsewhere.

similar shopping patterns leading to sizable intra-segment correlations, a cluster (area segment) of size one (i.e., one dot on the Master Sample map) was designated as the sampling unit. The reason for this becomes intuitively obvious when one considers that under this assumption, additional farms in the same area would tend only to duplicate the information obtained from a single farm.

In order to achieve the desired number of farm households (600), some adjustment of the Master Sample material was necessary in order to compensate for the decrease in the number of farms since the construction of the sample. In this study, the adjustment consisted of increasing the number of segments to be taken in a given county by the factor:

$$\frac{\text{number of Master Sample farms in the county}}{\text{number of farms in county according to the 1959 Census of Agriculture}}$$

For example, as a result of proportional allocation, six farms were to be taken from Allamakee County. The Master Sample indicated 2,362 farms (dots) in the open country zone, while the 1959 Census of Agriculture reported 1,717 farms. Thus, 1.375 Master Sample dots were equivalent to one farm in 1959. Consequently, the sample in Allamakee County consisted of eight segments (1.375 x 6) which were expected to yield six farms. Over the entire state, 743 such segments were drawn with the expectation that they would yield 600[3] farms. Within each county, segments were drawn with replacement. The sample was self-weighting, with a uniform sampling rate of approximately 1 in 291.[4] Although technically the counties corresponded to strata, the small sample size per county necessitated that counties be grouped for purposes of analysis; therefore, the six census economic areas were designated as strata.

TRAINING OF INTERVIEWERS

A two-day training school was conducted to instruct the interviewers in all phases of their work. A manual describing the procedures to be followed, including detailed instructions on the questionnaires, was prepared for their use as a guide during the training school and as a reference during the subsequent field work. Practice interviews were conducted during the training school. In addition, every interviewer was observed by a supervisor in at least one actual interviewing situation at the beginning

---

[3] For reasons discussed elsewhere, the number of farms actually found was considerably short of the desired 600.

[4] Slight fluctuations occur from county to county because of rounding the number of segments to integers. Also, this basic sampling rate does not include any adjustment for nonresponse, subsampling, etc.

of his assignment. Periodic supervision in the field was carried on throughout the field work phase of the study.

GENERAL FIELD PROCEDURE

As was stated previously, the sampling unit was an area segment. All households in the segments which were outlined on county maps were to be included regardless of whether or not they were represented by a dot on the map. The interviewer was to sketch each segment as he canvassed it, marking the location of each household with a household identification number. Vacant dwellings and segments containing no dwellings were identified by appropriate notation rather than merely by the absence of any household identification.

A questionnaire pertaining to the household was to be completed for each household in the sample. If the household contained a farm operator,[5] an additional questionnaire pertaining to the farm business was completed. If the household contained more than one farm operator or if an operator had more than one distinct operation, separate farm questionnaires were completed for each.

SPECIAL SITUATIONS IN FIELD PROCEDURES

Although the survey was conducted in the spring of 1961, information was sought for all of 1960. Since the population was not static, special procedures were adopted for situations in which changes had occurred between January 1, 1960, and the interview date. For example, persons living in a house in the segment at the time of enumeration who had moved there after March 1, 1960, were included in the sample only if they had lived somewhere else in the open country zone previous to the change of residence. The data were collected for the entire year just as if these people had been in the same location. Persons moving into the open country zone from a town or city after March 1 (hereafter referred to as ineligible households) were not included in the sample, since the nature of the information sought precluded any interest in persons who had been living in a town or city for any substantial part of 1960. On the other hand, persons who in 1960 had lived in a dwelling included in the present sample but had moved away prior to the interviewer's visit were not, in general, traced down and interviewed. Those moving elsewhere in the open country still had a chance to be included in the sample (see above); those moving into a town or city were essentially lost from the universe.[6]

---

[5] This term is defined elsewhere.

[6] Actually, as will be discussed later, some of these persons were traced down. In general, however, the cost of such an operation is prohibitive relative to the gain.

Since the Master Sample materials were prepared, many areas in the open country zone around urban centers have been transformed into housing developments and thus contain far more households than are indicated on the Master Sample maps. In this study, three of the sample segments fell into areas of this type. In order to avoid the considerable expense of interviewing all the households in these segments, a subsampling procedure was employed by which a known fraction of the households was interviewed.

After completion of most of the field work, 40 segments were found to contain households for which questionnaires were not completed because of various reasons.[7] Substitute segments were drawn to replace these households. Out of the 40 substitute segments, 2 contained no households and 6 contained households for which, again, questionnaires were not obtainable. Thus the apparent nonresponse rate was substantially reduced.

One hundred seventy segments were found to be vacant in the initial canvass. As a check on the quality of the field work, a sample of approximately one-half these segments was selected for revisit. Five additional farm households were found in this check.

Twenty-one segments were found to contain only ineligible nonfarm households (i.e., households whose occupants had moved into the open country zone after March 1, 1960). Fourteen of these segments were revisited in order to determine whether or not the previous occupant had been a farm operator at this place in 1960 and had moved out of the open country zone (thus having no chance of being enumerated in 1961). If this were the case, the interviewer located this person and completed the necessary questionnaires. Three additional farms were added to the sample by this procedure.

DEFINITIONS

Dwelling unit (1950 Census definitions)

In general, a dwelling unit is a group of rooms or a single room occupied (or intended for occupancy) as separate living quarters by a family or other group of persons living together or by a person living alone. Specifically, the above constitutes a dwelling unit if it has either 1) separate cooking equipment, or 2) two or more rooms and a separate entrance. Houses, apartments or flats, trailer houses, and living quarters above or in back of places of business are common examples of dwelling units.

Household

A household consists of those persons residing in a dwelling unit. Thus,

[7] Thirty questionnaires were not completed because of refusals, 5 in which the household was an ineligible farm household, and 5 for miscellaneous reasons.

there is a one-to-one correspondence between dwelling units and households, and the terms are often used synonymously.

Farm (general definition)

A farm consists of all the tracts of land, contiguous or noncontiguous, under the operation of a single individual or a group of individuals. An operator usually owns at least part of the assets but, as in the case of a hired manager, he need not. The farm acreage includes woodland, pasture, wasteland, etc., as well as cultivated land. In addition to the type of operation usually thought of as a farm, special operations such as apiaries, greenhouses and nurseries, feed lots, etc., are considered to be farms.

Farm (1959 Census definition)

In order to qualify as a census farm, places such as those just described must meet the following conditions:

1. If the place is less than ten acres in size, at least $250 worth of agricultural products must have been sold from the place in 1960 (of which at least $125 must have come from something other than forest products).
2. If the place is ten or more acres in size, at least $50 worth of agricultural products must have been sold from the place in 1960 (of which at least $25 must have come from something other than forest products).

Farm operator

A farm operator is a person actively engaged in running a farm. He must participate in the decision-making function and supply at least part of the labor.

Partnership

A partnership is a joint operation of a farm by two or more persons. These persons need not have a written agreement nor need they be related. In this study, a person in order to be considered a partner had to 1) work on the place at least 90 days in 1960, 2) share in the decision-making, and 3) receive a share of the profits (or absorb a share of the loss).

Principal partner

In this study, the junior partner (i.e., the younger or youngest) was considered the principal partner. The partnership operation entered the sample only with the principal partner. Consequently, if a junior partner lived in the segment, both household and farm questionnaires were completed; if a senior partner lived in the segment, only a household questionnaire was completed.

Hired manager

A hired manager does not usually own any land or capital in the farm he operates. He is considered to be an operator because he is hired to make the decisions and is in direct control of the operation.

Homemaker

The homemaker is the person who manages the home. Ordinarily the homemaker will be the wife of the operator, but this need not be the case. The homemaker may be a daughter, a sister, or a mother of the operator or she may be a hired housekeeper. In some cases, the operator himself may also be the homemaker.

COMPARISON OF NUMBER OF FARMS IN SAMPLE WITH NUMBER EXPECTED

As was stated earlier, the original expectation was 600 farms. However, this expectation was based on the total number of farms in the state in 1959 and was erroneously high. When the census figures are adjusted to the universe sampled (the open country zone) and are reduced to reflect one year's losses in number of farms, the expectation is reduced to 556 farms. The sample yielded a total of 530 farms. Of this total, 497 were interviewed and 21 were contacted but not interviewed (refusals, etc.). An additional 12 farms were added as adjustments resulting from the subsampling in built-up segments (5 farms), the check of a subsample of segments originally classed as vacant (5 farms), and the check of a subsample of segments containing only nonfarm, ineligible households (2 farms). In the latter operation, when it was discovered that the previous occupant had operated the place during the 1960 crop season and had since moved out of the open country zone, he (rather than the present occupant) was considered to be in the sample (cf. footnote 6, Chapter VI).

An approximate 95 per cent confidence interval placed around the sample number has an upper limit of 551, indicating that the discrepancy is slightly outside the sampling error. However, it must be remembered that the presample expectation is based on approximations, the accuracy of which cannot be verified. The adjustment to the open country zone is based on work by the late Margaret Haygood of the United States Department of Agriculture. Since this work was done over 15 years ago, the degree to which her findings reflect the present situation cannot be determined. At that time, she found that approximately 94 per cent of the farms in Iowa had their headquarters (residence of operator) in the Master Sample open country zone. The adjustment for losses in number of farms from 1959 to 1960 (1½ per cent) is based on the results of another survey conducted by the Iowa State University Statistical Laboratory and is, of course, subject to sampling error. The purpose of these presample adjustments is to obtain some idea of the sampling rate necessary to yield a predetermined number of farms. Ordinarily, differences between the presample estimates and estimates based on the sample data are ascribed to inaccuracies in the former.

ESTIMATION OF POPULATION MEANS AND VARIANCES

Since an approximately uniform sampling fraction was used, population means were easily estimated by the simple sample mean. Furthermore, since the segments were so small, the clustering that did occur can be ignored and estimates of the variance computed using the formula for stratified random sampling.

Let

y_{hij} = observation on j^{th} unit, i^{th} segment, h^{th} stratum where strata are defined as census economic areas

n_{hi} = number of units, i^{th} segment, h^{th} stratum.

Estimates of population means are obtained by

$$\hat{\bar{Y}} = \bar{y} = \sum_{h}\sum_{i}\sum_{j} y_{hij} \Big/ \sum_{h}\sum_{i} n_{hi} = \frac{1}{n_{..}} \sum_{h}\sum_{i}\sum_{j} y_{hij}$$

Ignoring the finite population correction, variances can be estimated by

$$v(\hat{\bar{Y}}) = \frac{1}{n^2} \sum_{h} \frac{n_{h.}}{n_{h.} - 1} \sum_{i}\sum_{j} (y_{hij} - \bar{y}_h)^2 .$$

$$\text{where } \bar{y}_h = \sum_{i}\sum_{j} y_{hij}/n_h$$

Use of the random sampling formula will tend to underestimate the variance. On the other hand, using the census economic areas rather than the individual counties as strata inflates the variance.

If estimates of state totals for farms are desired, they can be obtained by

$$\hat{Y} = 161{,}711 \; \hat{\bar{Y}}$$

where 161,711 is the estimated total number of farms in the open country zone of Iowa in 1960 and $\bar{Y}$ is defined as above. The variance can be estimated by

$$v(\hat{Y}) = (161{,}711)^2 \; v(\hat{\bar{Y}})$$

APPENDIX B

Description of Major Computer Programs Devised for the Analysis

The following is a description of the major programs written specifically to perform the analyses described in Chapters 4 and 5. All are written in Fortran IV, and after a description of their purpose, an annotated program listing of each is provided.

Program: LOCTYPE

This program is designed to compute for each member of a sample choosing from a set of alternatives, the types of alternatives available to it and the types it chooses. Each alternative belongs to a type defined in terms of attributes describing the alternative and its relationship to the sample member.

In this case, the sample members are households choosing from alternative towns which are classified into types - location types - in terms of population and distance to household.

The basic information required is a definition of each type. Also for each sample member, data on the alternatives chosen are required.

In this case, location types are defined in terms of population/ distance to household categories. Towns are described by their population and by their co-ordinate location, which in conjunction with the co-ordinate location of households enables household-to-town distances to be computed. Given that data, together with the towns chosen by each household, each household's alternatives, both chosen and rejected, can be assigned to location types. A sample of the program's output is shown in Table 4.1.

```
      PROGRAM LOCTYPE(INPUT,OUTPUT)
      DIMENSION ITD(1342),IE(1342),IN(1342),IPOP(1342),IDVIS(10),IPERCEN
     1 (10),ILIMP(7),ILIMD(7,5),IPOSS(7,5),IVIS(7,5,10),LOCNM(34),LOCVIS
     2(10)
C  READ IN DATA DESCRIBING EACH TOWN≠S POPULATION(IPOP),EASTING(IE),
C  AND NORTHING(IN).
      READ 1,(ITD(J),IE(J),IN(J),IPOP(J),J=1,1342)
C  READ IN THE DISTANCE AND POPULATION LIMITS(ILIMD,ILIMP) OF EACH
C  POSSIBLE LOCATION TYPE.
      READ 4,(ILIMP(J),J=1,7)
      READ 5,(ILIMD(1,K),K=1,3)
      READ 6,((ILIMD(J,K),K=1,4),J=2,7)
      PRINT 600
      PRINT 92
      PRINT 93
      PRINT 499
C  READ IN A HOUSEHOLD≠S EASTING(IHE),NORTHING(IHN),TOWNS PATRONISED
C  (IDVIS),THE PERCENTAGE OF GROCERY DOLLAR EXPENDITURE IN EACH(IPERCEN)
C  AND THE NUMBER PATRONISED(JB).
  832 READ 2,MD,JB,JKM,IHD,IHE,IHN,(IDVIS(J),IPERCEN (J),J=1,JB)
      IF(IHD.EQ.7)GO TO 832
      IF(MD.EQ.99) GO TO 1000
C  INITIALISE MATRICES.
      DO 300 K=1,7
      DO 300 J=1,5
      DO 301 L=1,10
  301 IVIS(K,J,L)=0
  300 IPOSS(K,J)=0
      DO 303LM=1,10
  303 LOCVIS(LM)=0
C  ORDER TOWNS PATRONISED ACCORDING TO DOLLAR EXPENDITURE.
      JA=0
      JL=JB-1
      DO 10J=1,JL
      JA=J+1
      DO 10JK=JA,JB
      IF(IPERCEN (J).LT.IPERCEN (JK))3,10
    3 JTEMP=IPERCEN (J)
      IPERCEN (J)=IPERCEN (JK)
      IPERCEN (JK)=JTEMP
      IDTEMP=IDVIS(J)
      IDVIS(J)=IDVIS(JK)
      IDVIS(JK)=IDTEMP
   10 CONTINUE
C  CALCULATE THE HOUSEHOLD≠S NUMBER OF ALTERNATIVES IN EACH LOCATION
C  TYPE(IPOSS).
      DO 20I=1,1342
C  CALCULATE THE MANHATTAN METRIC DISTANCE BETWEEN THE HOUSEHOLD AND
C  TOWN(IDIST).
      IDIST=IABS(IHE-IE(I))+IABS(IHN-IN(I))
      DO 20K=1,7
      IF(IPOP(I).LT.ILIMP(K))21,20
   21 IF(K.EQ.1)22,24
   22 DO 23J=1,3
      IF(IDIST.LT.ILIMD(K,J))222,23
   23 CONTINUE
C  CHECK IF ANY TOWN PATRONISED LIES OUTSIDE THE LIMITS OF ALL LOCATION
```

```
C   TYPES.
 84 DO 86LS=1,JB
    IF(ITD(I).EQ.IDVIS(LS))85,86
 85 PRINT 29,ITD(I),IDIST,IPOP(I),IHD
    GO TO 200
 86 CONTINUE
    GO TO 200
222 IPOSS(K,J)=IPOSS(K,J)+1
    DO 25L=1,JB
    IF(ITD(I).EQ.IDVIS(L))26,25
 26 IVIS(K,J,L)=IVIS(K,J,L)+1
    IPOSS(K,J)=IPOSS(K,J)-1
    GO TO 200
 25 CONTINUE
    GO TO 200
 24 DO 27J=1,4
    IF(IDIST.LT.ILIMD(K,J))222,27
 27 CONTINUE
    GO TO 84
 20 CONTINUE
200 CONTINUE
C   CONVERT THE MATRIX SHOWING THE NUMBER OF ALTERNATIVES IN EACH
C   LOCATION TYPE(IPOSS) INTO A SINGLE ARRAY(LOCNM).
    LM=0
    DO 400KK=1,3
    LM=LM+1
400 LOCNM(LM)=IPOSS(1,KK)
    DO 401JJ=2,7
    DO 401KK=1,4
    LM=LM+1
401 LOCNM(LM)=IPOSS(JJ,KK)
C   CONVERT THE MATRIX OF LOCATION TYPES PATRONISED(IVIS) INTO A SINGLE
C   ARRAY OF ORDERED LOCATION TYPES(LOCVIS).
    DO 402KJ=1,JB
    DO 402KL=1,3
    IF(IVIS(1,KL,KJ).NE.0)403,402
403 LOCVIS(KJ)=KL
402 CONTINUE
    DO 404KJ=1,JB
    DO 404LJ=2,7
    DO 404KL=1,4
    IF(IVIS(LJ,KL,KJ).NE.0)405,404
405 LOCVIS(KJ)=3+((LJ-2)*4)+KL
404 CONTINUE
C   WRITE OUT THE HOUSEHOLD≠S EASTING(IE),NORTHING(IN),NUMBER OF TOWNS
C   PATRONISED(JB),NUMBER OF ALTERNATIVES IN EACH LOCATION TYPE(LOCNM),
C   AND THE LOCATION TYPE OF EACH TOWN PATRONISED IN ORDER OF DOLLAR
C   EXPENDITURE(LOCVIS).
    PRINT 500,            IHD,IHE,IHN,(LOCNM(J),J=1,27),(LOCVIS(K),IPERC
   1EN(K),K=1,JB)
    GO TO 832
  1 FORMAT(5(I4,2I3,I6))
  2 FORMAT(3I2,3I4,(3(I5,6X,I7)))
  4 FORMAT(7I8)
  5 FORMAT(3I3)
  6 FORMAT(4I3)
 29 FORMAT(* WARNING NO DISTANCE CATEGORY FOR TOWN ID*I5*DISTANCE*I6*P
   1OPULATION*I7*HOUSEHOLD ID*I6)
```

```
92 FORMAT(1H0,32X,*LOCATION TYPES AVAILABLE*46X,*LOCATION TYPES VISIT
  1ED AND*)
93 FORMAT(* IHD IHE IHN       1  2  3  4  5  6  7  8  9 10 11 12 13 1
  14 15 16 17 18 19 20 21 22 23 24 25 26 27      PERCENTAGE OF $ SPENT
  2 IN EACH*)
499 FORMAT(1H0)
500 FORMAT(1X,I3,2I4,5X,27I3,5X,(6(I2,I3)))
600 FORMAT( 1H0,*LOCATION TYPES AVAILABLE TO EACH HOUSEHOLD AND THE  P
  1ROPORTIONATE DOLLAR EXPENDITURE IN EACH TYPE VISITED*)
1000 STOP
     END
```

Program: PCOMP

This program is designed to calculate the implicit paired comparisons made by each household in visiting some location types and rejecting other available ones. It is written separately from LOCTYPE so that different aspects of the household's choices may be analysed with simple alterations to this program, but without requiring to regenerate the information provided by LOCTYPE. Thus, the paired comparisons implicit in all of a household's choices or only in its first choice may be obtained using only this program.

```
      PROGRAM PCOMP  (INPUT,OUTPUT,TAPE1,TAPE2)
      DIMENSION LOCPOS(34),LOCVIS(10),PROP(34,34),IPERCEN(10),F(34,34),
     1AVPROP(34,34),IWIN(400),ILOSE(400),PROX(400),I(34,34)
      DO 56K=1,JZ
      DO 56KN=1,K
   56 I(K,KN)=0
      PRINT 600
  600 FORMAT(1H0,*PAIRED COMPARISONS MADE BY EACH HOUSEHOLD DERIVED FROM
     1 THE PROPORTIONATE DOLLAR EXPENDITURE IN EACH LOCATION TYPE AVAILA
     2BLE*)
C  READ IN THE HOUSEHOLD≠S ID(IHD),THE NUMBER OF TOWNS IT PATRONISED(JB)
C  ,THE NUMBER OF TOWNS AVAILABLE TO IT IN EACH LOCATION TYPE(LOCPOS),
C  THE LOCATION TYPES PATRONISED(LOCVIS) IN ORDER OF PERCENTAGE GROCERY
C  DOLLAR EXPENDITURE(IPERCEN).
   18 READ(1,1)NUM,JB,(LOCPOS(N),N=1,JZ),(LOCVIS(M),M=1,JB)
    1 FORMAT(1X,I2,6X,I3,2X,27I2,2X,(10I3))
      IF(EOF,1)1101,1102
C  INITIALISE MATRICES.
 1102 DO 10J=1,JZ
      DO 10L=1,J
      F(J,L)=0.
      AVPROP(J,L)=0.
   10 PROP(J,L)=0.
      IP=0
C  FOR EACH LOCATION TYPE PATRONISED BY THE HOUSEHOLD, FIND ALL OTHERS
C  PATRONISED AND REJECTED TO WHICH IT WAS IMPLICITLY PREFERRED.
      DO 1000K=1,JB
      IF(LOCVIS(K).EQ.0)1000,19
   19 DO 20J=1,JZ
      IF(LOCVIS(K).EQ.J)50,20
   20 CONTINUE
   50 IF(K.EQ.JB)GO TO 100
      JA=K+1
```

```
      DO 180MM=JA,JB
      IF(LOCVIS(MM).EQ.0)GO TO 180
      DO 80KJ=1,JZ
      IKJ=KJ
      IF(LOCVIS(MM).EQ.KJ)91,80
   80 CONTINUE
   91 IF(IKJ.LE.J)92,93
C  THE PROPORTIONATE PREFERENCE OF ONE PATRONISED LOCATION TYPE OVER
C  ANOTHER PATRONISED TYPE(PROP) EQUALS THE PERCENTAGE SPENT IN THE
C  MOREPREFERRED DIVIDED BY THE PERCENTAGE SPENT IN BOTH.
   92 PROP(J,IKJ)=(FLOAT(IPERCEN(K))/(FLOAT(IPERCEN(K)+IPERCEN(MM))))+
     1PROP(J,IKJ)
      F(J,IKJ)=F(J,IKJ)+1.
      I(J,IKJ)=I(J,IKJ)+1
      GO TO 180
   93 PROP(IKJ,J)=1.-(FLOAT(IPERCEN(K))/(FLOAT(IPERCEN(K)+IPERCEN(MM))))
     1+PROP(IKJ,J)
      F(J,IKJ)=F(J,IKJ)+1.
      I(J,IKJ)=I(J,IKJ)+1
  180 CONTINUE
  100 DO 110LA=1,JZ
      IF(LOCPOS(LA).EQ.0)110,120
  120 IF(LA.LE.J)121,122
C  THE PROPORTIONATE PREFERENCE OF A PATRONISED LOCATION TYPE OVER A
C  REJECTED TYPE(PROP) IS ONE.
  121 PROP(J,LA)=PROP(J,LA)+1.
      F(J,LA)=F(J,LA)+1.
      I(J,LA)=I(J,LA)+1
      GO TO 110
  122 F(LA,J)=F(LA,J)+1.
      I(LA,J)=I(LA,J)+1
  110 CONTINUE
 1000 CONTINUE
      DO 123KP=1,JZ
      DO 123JP=1,KP
  123 AVPROP(KP,JP)=PROP(KP,JP)/F(KP,JP)
      DO 200LL=1,JZ
      DO 200KK=1,LL
      IF(PROP(LL,KK).EQ.0..AND.F(LL,KK).EQ.0.)200,201
  201 IP=IP+1
      IWIN(IP)=LL
      ILOSE(IP)=KK
      PROX(IP)=AVPROP(LL,KK)
  200 CONTINUE
      IF(IP.EQ.0) GO TO 18
C  WRITE OUT THE NUMBER OF TOWNS PATRONISED BY THE HOUSEHOLD, THE MORE
C  PREFERRED(IWIN),AND LESS PREFERRED(ILOSE) LOCATION TYPES ,AND THE
C  PREFERENCE AMOUNT(PROX) FOR ALL IMPLICIT PAIRED COMPARISONS BY THE
C  HOUSEHOLD.
      PRINT 203,IHD,IP,(IWIN(K),ILOSE(K),PROX(K),K=1,IP)
  203 FORMAT(*0*2I3,(12(I3,I2,F5.2)))
      GO TO 18
 1101 ENDFILE 2
      STOP
      END
```

Program: Paired Comparison Preference Analysis (PCPA)

This program uses the paired comparison data for each household in a sample, such as program PCOMP provides, and analyses the aggregate preference structures of groups and the total sample, derived from the individual data. The following major operations are performed:

1. Aggregate the individual paired comparisons of any set of households in the form of a paired comparison matrix showing the proportion of times each location type is preferred to each other.
2. Recompute the set's paired comparison matrix using an additivity rule to compute indirect estimates of the amount one type is preferred to another, in addition to the original observed proportion of times the one is preferred to the other.
3. Compare the preference matrices of subsets of the sample using statistical tests of the null hypothesis that $p_{ij1} = p_{ij2}$ for all cells in the matrix.

```
      PROGRAM PCPA(INPUT,OUTPUT,PUNCH,TAPE4,TAPE5=INPUT,TAPE6=OUTPUT)
      DIMENSION IBAR(30),IWIN(60),ILOSE(60),PROX(60),IDAT(20),IVARR(20),
     1SUML(500),SUMU(500),           LOCTYPL(30),LOCTYPU(30)    ,F(5),
     2NSUB(27,25),PROP(27,25),IBAR1(30),BAR(5),BAR1(5),SUMSUB(30),SUMCT(
     330),AVSUM(30)
      COMMON        NL(500),NU(500),SUMONEL(500),SUMONEU(500),AGMATL(500),
     1AGMATU(500),Z1(500),TYPTEST(500),DIFGRO(27,27),FDIFF(500),
     2NCOUNT(27,25),LB,JROW,KCOL,LVAR,CUMPROB(30),IY,JZ,JMN,PROXL(500),
     3PROXU(500),OP(40),OPPUN(4,30),NDIS(60,3)
      REAL NL,NU,NCOUNT,NSUB,L1,L2,L1E,L2E
C  INITIALISE THE UNIFORM RANDOM NUMBER GENERATOR.
      CALL RANDU(7,IY,Y)
      DO150 LRR=1,100
      CALL RANDU(IY,IY,Y)
  150 CONTINUE
      REWIND 4
C  READ IN OPTION PARAMETERS INDICATING A)THE FIRST PAIR OF GROUPS TO BE
C  COMPARED(JYY), B)THE LAST TO BE COMPARED(JY), C)THE NUMBER OF
C  LOCATION TYPES(JZ),AND D) IF JADD=1,A TEST FOR ADDITIVITY IS REQUIRED
      READ(5,12)JYY,JY,JZ,JMN,JADD
   12 FORMAT(5I5)
C  READ IN THE UPPER VARIABLE SCORE LIMITS FOR LOWER GROUPS.
      READ(5,2)(IBAR (J),J=1,25)
C  READ IN THE LOWER VARIABLE SCORE LIMITS FOR UPPER GROUPS.
      READ(5,2)(IBAR1(J),J=1,25)
    2 FORMAT(20I4)
      WRITE(6,503)(IBAR(J),J=1,25)
      WRITE(6,503)(IBAR1(J),J=1,25)
  503 FORMAT(1H ,30I4)
C  READ IN THE UPPER FACTOR SCORE LIMITS FOR LOWER GROUPS.
      READ(5,502)(BAR(J),J=1,5)
C  READ IN THE LOWER FACTOR SCORE LIMITS FOR UPPER GROUPS.
      READ(5,502)(BAR1(J),J=1,5)
  502 FORMAT(10F5.2)
      READ(5,405)IND,(IDAT(J),J=1,IND)
  405 FORMAT(40I2)
C  READ IN OUTPUT LIST OPTIONS, AND THE GROUPINGS WITH WHICH SUBROUTINE
C  RECOMP IS TO BE USED.
      READ 532,(OP(J),J=1,30)
  532 FORMAT(30F2.0)
      WRITE(6,504)(OP(J),J=1,30)
  504 FORMAT(1X,30F3.0)
      READ 532,((OPPUN(J,K),K=1,30),J=1,4)
      WRITE(6,504)((OPPUN(J,K),K=1,30),J=1,4)
      JZZ=JZ-1
      LVAR=0
      JT=0
      KQ=0
      JTT=0
      JN=0
C  INITIALISE ARRAYS AND MATRICES.
      DO 9K=1,JZ
      DO 7J=1,25
      PROP(K,J)=0.
      NSUB(K,J)=0.
    7 NCOUNT(K,J)=0.
      DO 9L=1,K
```

```
      KL=((K*(K-1))/2)+L
      SUML(KL   )=0.
      SUMONEL(KL   )=0.
      NL(KL   )=0.
      SUMU(KL   )=0.
      SUMONEU(KL   )=0.
    9 NU(KL   )=0.
      DO 901 J=1,30
      SUMSUB(J)=0.
      SUMCT(J)=0.
      AVSUM(J)=0.
  901 CONTINUE
      DO 764 J=1,60
      DO 764 L=1,3
      NDIS(J,L)=0
  764 CONTINUE
  505 IF(OP(14).EQ.1.)3,601
C  READ IN ONE HOUSEHOLD ID(ID) AND DATA ON ALL ITS PAIRED COMPARISONS
C  WHERE LOCATION TYPE IWIN IS PREFERRED TO TYPE ILOSE BY A DEFINED
C  PROPORTION(PROX). IP=THE NUMBER OF PAIRED COMPARISONS BY THE
C  HOUSEHOLD.
C      READ TAPE NO.ZZ1416  ALL TOWN CHOICES PAIRED COMPARISONS
C      NOTE  FOR 1ST TOWN CHOICE PAIRED COMPARISON ALTER READ FORMAT TO
C  517 AND ALTER REQUEST TO REQUEST,TAPE3,HY.  REEL NO ZZ1416 TO READ
C  ONLY. ALSO ADD REWIND(TAPE4) AND COPYCF(TAPE3,D,1) AND COPYCF
C  (TAPE3,TAPE4,1) TO CONTROL CARDS BEFORE RUN(S).
C
    3 READ(4,517)ID, IP,(IWIN(K),ILOSE(K),PROX(K),K=1,IP)
  570 IF(EOF,4)102,566
  566 JN=JN+1
      DO 930J=1,IP
C  COMPUTE 2 PAIRED COMPARISON MATRICES FOR THE TOTAL SAMPLE. THE FIRST
C  INDICATES PROPORTIONATE PREFERENCE USING INFORMATION ON PERCENTAGE
C  EXPENDITURE IN LOCATION TYPES. THE SECOND INDICATES PROPORTIONATE
C  PREFERENCE USING ONLY IMFORMATION ON WHICH OF EACH PAIR OF LOCATION
C  TYPES A HOUSEHOLD PREFERS.
      L=IWIN(J)
      LL=ILOSE(J)
   31 KK=((L*(L-1))/2)+LL
C  IN CALCULATING THE FIRST PREFERENCE MATRIX, FOR EACH PAIRED
C  COMPARISON BY EACH HOUSEHOLD, ADD PROX TO THE CELL WHICH INDICATES
C  LOCATION TYPE IWIN IS PREFERRED TO ILOSE.
      SUML   (KK)=SUML   (KK)+PROX(J)
      NL   (KK)=NL   (KK)+1.
C  IN CALCULATING THE SECOND PREFERENCE MATRIX, FOR EACH PAIRED
C  COMPARISON BY EACH HOUSEHOLD, ADD 1 TO THE CELL WHICH INDICATES
C  LOCATION TYPE IWIN IS PREFERRED TO ILOSE.
      IF(PROX(J).GT.0.50)SUMONEL(KK)=SUMONEL(KK)+1.
  930 CONTINUE
      GO TO 3
  102 REWIND 4
      PRINT 501,JN
  501 FORMAT(*0    NUMBER OF HOUSEHOLDS WITH DATA ON BEHAVIOR=*I4)
  300 PRINT 45
   45 FORMAT(1H1,* $$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
     1$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
     2*/1H ,45X,*  LOCATIONAL PREFERENCES OF TOTAL SAMPLE*/1H ,* $$$$$$
     3$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
```

```
    4$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*)
      PRINT 17,JN
   17 FORMAT(*0    PPEFERENCES FOR ROW LOCATION TYPES AGAINST COLUMN LOCA
     1TION TYPES AS REPRESENTED BY PROPORTIONATE ALLOCATION OF DOLLARS*/
     2*0    BETWEEN EACH PAIR OF LOCATION TYPES*/*0    1. INDICATES TOTAL
     3EXPENDITURE IN ROW LOCATION TYPE*/*0    0. INDICATES TOTAL EXPENDIT
     4URE IN COLUMN LOCATION TYPE */*0    -9.9 INDICATES NO DATA*/*0    VA
     5LUES CALCULATED FROM ALL GROCERY EXPENDITURES BY EACH HOUSEHOLD,N=
     6*I3)
C  CALL THE SUBROUTINE MATCALC WHICH CALCULATES THE AVERAGE
C  PROPORTIONATE PREFERENCE BETWEEN EACH PAIR OF LOCATION TYPES USING
C  BOTH OF THE ABOVE SETS OF INFORMATION AND WRITE OUT THE 2 PREFERENCE
C  MATRICES.
      CALL MATCALC(SUML,SUMONEL,AGMATL,NL,JZ,OP,OPPUN(2,1))
  202 IF(OPPUN(4,1).EQ.0.)GO TO 203
C  CALL THE SUBROUTINE RECOMP WHICH RECOMPUTES AGGREGATE PREFERENCE
C  MATRICES USING AN ADDITIVITY RULE TO COMPUTE ADDITIONAL INDIRECT
C  ESTIMATES OF THE AMOUNT ONE TYPE IS PREFERRED TO ANOTHER.
      CALL RECOMP(1)
C  CALL THE SUBROUTINE REORDER WHICH REARRANGES ROWS AND COLUMNS OF A
C  PREFERENCE MATRIX SO THAT LOCATION TYPES ARE RANKED IN ORDER OF
C  PREFERENCE FROM TOP TO BOTTOM, AND LEFT TO RIGHT.
  203 CALL REOPDER(AGMATL,NL,JZ,LOCTYPL,OP,OPPUN(3,1))
C  COMPUTE AND COMPARE PREFERENCE MATRICES FOR GROUPS DEFINED IN TERMS
C  OF VARIABLE SCORES, FOR ALL VARIABLES FROM THE JYTH TO THE JYYTH.
  601 DO 43K=JYY,JY
      REWIND 4
      LVAR=K
      MN=K+1
      DO 301J=1,JZ
      DO 301L=1,J
      KL=((J*(J-1))/2)+L
      SUML(KL)=0.
      SUMONEL(KL)=0.
      NL(KL)=0.
      SUMU(KL)=0.
      SUMONEU(KL)=0.
      NU(KL)=0.
      Z1(KL)=0.
      FDIFF(KL)=0.
      TYPTEST(KL)=1RX
      JNGL=0
  301 JNGU=0
      IF(OP(16).GT.0.)GO TO 315
C  READ LOGICAL UNIT NO.4.
C
C  READ IN ONE HOUSEHOLD ID, DATA ON ALL ITS PAIRED COMPARISONS, AND
C  ITS SCORES ON 25 VARIABLES(IVARR,F). IP=THE NUMBER OF PAIRED
C  COMPARISONS BY THE HOUSEHOLD.
  100 READ(4,515)ID,JB,IP,(IVARR(L),L=1,20),(F(J),J=1,5),(IWIN(N),ILOSE(
     1N),PROX(N),N=1,IP)
  515 FORMAT(I3,I2,10I2,I4,6I2,I4,3I2,5F6.3,6(2I2,F5.2)/(15(2I2,F5.2)))
      IF(EOF,4)403,310
C  ASSIGN THE HOUSEHOLD TO A RANDOM GROUP.
  310 IF(OP(16).LT.0.)313,314
  313 CALL RANDU(IY,IY,U)
      IF(U.LT..5)400,181
C  ASSIGN THE HOUSEHOLD TO ONE OF TWO GROUPS IF ITS KTH VARIABLE SCORE
```

```
C   (IVARR(K) OR F(KZ)) LIES WITHIN THE LIMITS DEFINED FOR ONE OF THE
C   GROUPS.
  314 IF(K.LE.20)404,311
  404 IF(IVARR(K).LE.IBAR(K))8,18
   18 IF(IVARR(K).GT.IBAR1(K))181,100
    8 IF(IVARR(K).GE.1)400,401
  401 DO 402J=1,IND
      IF(IVARR(K).LT.1.AND.K.EQ.IDAT(J))400,402
  402 CONTINUE
      GO TO 100
  311 KZ=K-20
      IF(F(1).EQ.0..AND.F(2).EQ.0..AND.F(3).EQ.0..AND.F(4).EQ.0..AND.F(5
     1).EQ.0.)100,312
  312 IF(F(KZ).LE.BAR(KZ))400,182
  182 IF(F(KZ).GT.BAR1(KZ))181,100
C       READ TAPE NO.AA0709 WITH HYPOTHETICAL BEHAVIOR
C   READ IN ONE HOUSEHOLD ID AND HYPOTHETICAL DATA ALL ITS PAIRED
C   COMPARISONS BASED ON A RANDOMLY ASSIGNED PREFERENCE RULE, DETERMINED
C   IN PROGRAM HYPBEH. THE HOUSEHOLD≠S LOCATION AND THE NUMBER OF TOWNS
C   PATRONISED ARE THE SAME AS FOR THE EMPIRICAL DATA.
  315 IF(OP(16).GT.1.)316,317
  316 READ(4,518)KZ,IP,(IWIN(L),ILOSE(L),PROX(L),L=1,IP)
  518 FORMAT(1X,I1,2X,I3,(1X,12(I3,I2,F5.2)))
      IF(EOF,4)403,407
C   ASSIGN THE HOUSEHOLD TO ONE OF TWO GROUPS ACCORDING TO WHICH OF 2 OR
C   MORE PREFERENCE RULES ITS BEHAVIOR CONFORMS TO.
  407 IF(KZ.GT.2)316,406
C   READ IN ONE HOUSEHOLD ID,AND SEPARATE SETS OF PAIRED COMPARISON
C   DATA BASED ON ITS 1ST CHOICE TOWN, AND ITS 2ND CHOICE TOWN, IF MORE
C   THAN ONE TOWN WAS CHOSEN.
  317 READ(5,516)ID,KZ,IP,(IWIN(N),ILOSE(N),PROX(N),N=1,IP)
  516 FORMAT(I3,2I2,1X,(8(2I2,F5.2)))
      IF(ID.GE.900)403,408
  408 IF(KZ.GT.2)317,406
C   ASSIGN HOUSEHOLD≠S PAIRED COMPARISONS BASED ON ITS 1ST CHOICE TOWN
C   TO ONE GROUP, AND ITS PAIRED COMPARISONS BASED ON ITS 2ND CHOICE TOWN
C   (IF ANY) TO THE OTHER GROUP.
  406 IF(KZ.EQ.1)400,181
  400 JNGL=JNGL+1
C   CALCULATE PREFERENCE MATRICES FOR BOTH GROUPS IN THE SAME MANNER AS
C   IS DESCRIBED ABOVE FOR THE TOTAL SAMPLE.
      DO910J=1,IP
      L=IWIN(J)
      LL=ILOSE(J)
   11 KQ=((L*(L-1))/2)+LL
      SUML(KQ  )=SUML(KQ  )+PROX(J)
      NL(KQ  )=NL(KQ  )+1.
      IF(PROX(J).GT.0.50)SUMONEL(KQ  )=SUMONEL(KQ  )+1.
  910 CONTINUE
      IF(OP(16).GT.0.)315,100
  181 JNGU=JNGU+1
      DO919J=1,IP
      L=IWIN(J)
      LL=ILOSE(J)
   22 KT=((L*(L-1))/2)+LL
      SUMU(KT  )=SUMU(KT  )+PROX(J)
      NU(KT  )=NU(KT  )+1.
      IF(PROX(J).GT.0.50)SUMONEU(KT  )=SUMONEU(KT  )+1.
```

```
  919 CONTINUE
      IF(OP(16).GT.0.)315,100
  403 PRINT 46,K,IBAR(K),JNGL
   46 FORMAT(1H1,*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
     1$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
     2/1H ,*LOCATIONAL PPEFERENCES OF HOUSEHOLDS WITH VALUE OF SOCIOECON
     3OMIC  VARIABLF*I3*LESS THAN OR EQUAL TO*I3*, N=*I3/
     4             1H ,*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
     5$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
     6$$$$$*)
      CALL MATCALC(SUML,SUMONEL,AGMATL,NL,JZ,OP,OPPUN(2,MN))
      PRINT 47,K,IBAR(K),JNGU
   47 FORMAT(1H0,*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
     1$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
     2/*   LOCATIONAL PREFERENCES OF HOUSEHOLDS WITH VALUE OF SOCIOECONO
     3MIC VARIABLE *I3* GREATER THAN*I3*,N=*I3/1H ,*$$$$$$$$$$$$$$$$$$$$
     4$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
     5$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*)
      CALL MATCALC(SUMU,SUMONEU,AGMATU,NU,JZ,OP,OPPUN(2,MN))
      IF(OPPUN(4,MN).EQ.0.)GO TO 204
      CALL RECOMP(MN)
  204 CALL PEORDER(AGMATL,NL,JZ,LOCTYPL,OP,OPPUN(3,MN))
      CALL REORDER(AGMATU,NU,JZ,LOCTYPU,OP,OPPUN(3,MN))
      PRINT 48,(LOCTYPL(J),J=1,JZ)
      PRINT 48,(LOCTYPU(J),J=1,JZ)
   48 FORMAT(1H0,30I4)
C   SUBTRACT THE PREFERENCE MATRICES OF THE 2 GROUPS.
      DO 71J=1,JZ
      DO 71JA=1,J
      I=((J*(J-1))/2)+JA
      DIFGRO(J,JA)=AGMATL(I)-AGMATU(I)
      IF(AGMATU(I).EQ.-9.9.OR.AGMATL(I).EQ.-9.9)DIFGRO(J,JA)=-9.9
   71 CONTINUE
      IF(OP(1).NE.1.)GO TO 74
      PRINT 72,K
   72 FORMAT(*0   DIFFERENCES BETWEEN PREFERENCES OF UPPER AND LOWER GRO
     1UPS AS DEFINED IN TERMS OF S E VARIABLE*I3/*0   (UPPER GROUPS PREF
     2ERENCE -LOWER GROUPS PREFERENCE)*/*0   -9.9 SIGNIFIES NO DATA*)
      PRINT 13
      DO 73L=1,JZ
   73 PRINT 42,L,(DIFGRO(L,JA),JA=1,L)
C  CONFIDENCE LEVEL FOR STATISTICAL TESTS IS ALPHA=.05.
   74 ALPHA=0.05
      IF(OP(2).NE.1.)GO TO 75
      IF(K.GT.3)GO TO 75
      PRINT 360,ALPHA
  360 FORMAT(*1              CHI2 STATISTICS AND FISHER EXACT PROBABILITY (T
     1OCHER S MODIFICATION) STATISTICS FOR TEST OF HYPOTHESIS*/*0 THAT P
     2 HAT(A,B) OF LOWER GROUP IS SIGNIFICANTLY DIFFERENT FROM P HAT(A,B
     3) OF UPPER GROUP, ALPHA=*F5.2)
      PRINT 361
  361 FORMAT(*0              1           2           3            4
     1     5           6           7           8           9          10
     2*)
   75 DO 420 JH=1,JZ
C  CHECK WHETHER THE P(IJ) CELLS IN BOTH MATRICES CAN BE COMPARED
C  STATISTICALLY. EXCLUDE FROM COMPARISON THOSE WHERE NO INFORMATION
C  IS AVAILABLE FOR EITHER GROUP, OR WHERE LOCATION TYPES 12,16,19 OR 20
```

```
C   ARE BEING COMPARED EITHER TO ANY OTHER LOCATION TYPE OR TO ONE
C   ANOTHER.
      IF(JH.EQ.12.OR.JH.EQ.16.OR.JH.EQ.19.OR.JH.EQ.20)GO TO 420
      DO 424 JD=1,JH
      IF(JH.EQ.JD)GO TO 424
      IF(JD.EQ.12.OR.JD.EQ.16.OR.JD.EQ.19.OR.JD.EQ.20)GO TO 424
      LB=((JH*(JH-1))/2)+JD
      JROW=JH
      KCOL=JD
      IF(AGMATL(LB).EQ.-9.9.OR.AGMATU(LB).EQ.-9.9)441,423
  441 Z1(LB)=-9.9
      FDIFF(LB)=-9.9
      NSUB(JROW,LVAR)=NSUB(JROW,LVAR)+1.
      NSUB(KCOL,LVAR)=NSUB(KCOL,LVAR)+1.
      GO TO 424
C   TEST WHETHER THE 2 P(IJ) VALUES AND THEIR ASSOCIATED SAMPLE SIZES
C   MEET THE REQUIREMENTS OF THE CHI SQUARED TEST. IF NOT USE THE
C   FISHER EXACT PROBABILITY TEST.
  423 IF((NL(LB)+NU(LB)).GE.40.)425,426
  425 CALL CHITEST
      GO TO 424
  426 IF((NL(LB)+NU(LB)).GE.20.)427,428
  427 W1=SUMONEL(LB)
      L1=NL(LB)-SUMONEL(LB)
      W2=SUMONEU(LB)
      L2=NU(LB)-SUMONEU(LB)
      W1E=((W1+W2)*NL(LB))/(NL(LB)+NU(LB))
      W2E=((W1+W2)*NU(LB))/(NL(LB)+NU(LB))
      L1E=((L1+L2)*NL(LB))/(NL(LB)+NU(LB))
      L2E=((L1+L2)*NU(LB))/(NL(LB)+NU(LB))
      IF(W1E.GE.5..AND.W2E.GE.5..AND.L1E.GE.5..AND.L2E.GE.5.)GO TO 425
  428 CALL FEPTEST
  424 CONTINUE
  420 CONTINUE
C   PRINT THE RESULTS OF EACH STATISTICAL TEST BETWEEN THE 2 PREFERENCE
C   MATRICES.
      IF(OP(2).NE.1.)GO TO 76
      IF(K.GT.3)GO TO 76
      DO 444L=1,JZ
      LA=((L*(L-1))/2)+1
      J=L
      IF(L.GT.10)J=10
      LB=((L*(L-1))/2)+J
      PRINT 431,L,(TYPTEST(LU),Z1(LU),LU=LA,LB)
  431 FORMAT(1H0,I3,10(R6,F6.2))
      PRINT 362,(DIFGRO(L,LR),LR=1,J)
  362 FORMAT(1H ,3X,10(6X,F6.2))
      PRINT 362,(FDIFF(LU),LU=LA,LB)
  444 CONTINUE
      PRINT 363
  363 FORMAT(*0                 11              12              13              14
     1     15              16              17              18              19              20
     2*)
      DO 364L=11,JZ
      LA=((L*(L-1))/2)+11
      J=L
      IF(L.GT.20)J=20
      LB=((L*(L-1))/2)+J
```

```
      PRINT 431,L,(TYPTEST(LU),Z1(LU),LU=LA,LB)
      PRINT 362,(DIFGRO(L,LR),LR=11,J)
      PRINT 362,(FDIFF(LU),LU=LA,LB)
  364 CONTINUE
      PRINT 365
  365 FORMAT(*0                    21                    22                    23                    24
     1      25                    26                    27*)
      DO 366L=21,JZ
      LA=((L*(L-1))/2)+21
      LB=((L*(L-1))/2)+L
      PRINT 431,L,(TYPTEST(LU),Z1(LU),LU=LA,LB)
      PRINT 362,(DIFGRO(L,LR),LR=21,L)
      PRINT 362,(FDIFF(LU),LU=LA,LB)
  366 CONTINUE
   76 JROW=JZ
      IF(OP(22).NE.1.)GO TO 43
      CALL DIRTEST
   43 CONTINUE
      IF(OP(3).NE.1.)STOP
      PRINT 887
  887 FORMAT(1H1)
C   PRINT FOR EACH PAIR OF COMPARED PREFERENCE MATRICES , THE NUMBER OF
C   TIMES EACH LOCATION TYPE WAS SIGNIFICANTLY DIFFERENTLY PREFERRED BY
C   THE 2 GROUPS.
      PRINT 888,((NCOUNT(J,K),K=1,25),J=1,JZ)
  888 FORMAT(1H ,5X,25F4.0)
      DO 896 K=1,25
      DO 897 J=1,JZ
      SUMSUB(K)=SUMSUB(K)+NSUB(J,K)
      SUMCT(K)=SUMCT(K)+NCOUNT(J,K)
  897 CONTINUE
      SUMCT(26)=SUMCT(26)+SUMCT(K)
      SUMSUB(26)=SUMSUB(26)+SUMSUB(K)
      AVSUM(K)=(SUMCT(K)/2.)/(253.-(SUMSUB(K)/2.))
  896 CONTINUE
      PRINT 888,(SUMCT(K),K=1,26)
      PRINT 889
  889 FORMAT(1H0)
  895 PRINT 888,((NSUB(J,K),K=1,25),J=1,JZ)
      PRINT 888,(SUMSUB(K),K=1,26)
      PRINT 889
      DO 890 J=1,JZ
      DO 891 K=1,25
      IF(NCOUNT(J,K).EQ.0..OR.NSUB(J,K).EQ.26.)891,893
  893 PROP(J,K)=NCOUNT(J,K)/(22.-NSUB(J,K))
  891 CONTINUE
      PRINT 894,(PROP(J,L),L=1,25)
  894 FORMAT(1H ,25F5.2)
  890 CONTINUE
      PRINT 894,(AVSUM(K),K=1,25)
      SS=0.
      DO 898 K=1,25
      SS=SS+AVSUM(K)
  898 CONTINUE
      AVAV=SS/25.
      PRINT 899,AVAV
  899 FORMAT(1H0,F6.3)
C   PRINT HISTOGRAMS OF CHI SQUARED AND FISHER EXACT PROBABILITY VALUES
```

```
C   CALCULATED IN ALL TESTS.
      DO 767 L=1,3
      DO 767 J=1,59
      NDIS(60,L)=NDIS(60,L)+NDIS(J,L)
  767 CONTINUE
      DO 765 L=1,3
      WRITE(6,766)(NDIS(J,L),J=1,60)
  766 FORMAT(1H0,30I4)
  765 CONTINUE
    4 FORMAT(I3,2X,I2,74X,6(2I2,F5.2)/(15(2I2,F5.2)))
  514 FORMAT(I3,2X,I2,9I2,I4,6I2,I4,3I2,5F6.3,6(2I2,F5.2)/(15(2I2,F5.2))
     1)
  517 FORMAT(I3,2X,I2,74X,                          6(2I2,F5.2)/(15(2I2,F5.2)))
   13 FORMAT(*0          1   2   3   4   5   6   7   8   9  10  11  12  13
     1 14  15  16  17  18  19  20  21  22  23  24  25  26  27*)
   42 FORMAT(* *I2,2X,27F4.1)
  119 FORMAT(1H0,4X,27I4)
  121 FORMAT(1H ,I2,2X,27F4.1,2X,F6.3,F4.0)
  192 FORMAT(* *I2,2X,27F4.0)
  232 FORMAT(1H ,I2,2X,27F4.0,4X,F4.0)
 9211 STOP
      END
```

```
      SUBROUTINE MATCALC(SUM,SUMONE,AGMAT,N,JZ,OP,OPP)
      DIMENSION SUM(500),SUMONE(500),N(500),GROMAT(500),AGMAT(500)    ,
     1OP(40)
      REAL N
      GAMMA=0.
      DENOM=0.
      IF(OP(4).NE.1.)GO TO 1
      PRINT 44
   44 FORMAT(*0   PREFERENCES FOR ROW LOCATION TYPES AGAINST COLUMN LOCA
     1TION TYPES*/*0   -9.9 SIGNIFIES NO DATA*)
      PRINT 13
    1 DO 45LA=1,JZ
      JT=((LA*(LA-1))/2)+1
      DO 46LB=1,LA
      KS=((LA*(LA-1))/2)+LB
      IF(LB.EQ.LA)30,31
   30 GROMAT(KS)=.5
      AGMAT(KS)=.5
      GO TO 46
   31 IF(N (KS   ).EQ.0.)170,171
  170 GROMAT (KS)=-9.9
      AGMAT (KS)=-9.9
      GO TO 46
  171 GROMAT (KS)=SUM (KS   )/N (KS   )
      AGMAT (KS)=SUMONE (KS   )/N (KS   )
C  CALCULATE THE DENOMINATOR AND NUMERATOR OF KENDALL≠S COEFFICIENT
C  OF AGREEMENT. THIS COEFFICIENT INDICATES THE PERCENTAGE OF AGREEMENT
C  BETWEEN A GROUP OF HOUSEHOLDS≠ PAIRED COMPARISONS.
      IF(N (KS   ).EQ.1.)GO TO 46
      IF(SUMONE (KS   ).EQ.0.)172,173
  172 GAMMA     =GAMMA     +((N (KS   )*(N (KS   )-1.))/2.)
      GO TO 174
  173 IF(SUMONE (KS   ).EQ.1..AND.N (KS   ).EQ.2.)174,175
  175 GAMMA     =GAMMA     +((SUMONE (KS   )*(SUMONE (KS   )-1.))/2.)
     1+(((N (KS   )-SUMONE (KS   ))*(N (KS   )-SUMONE (KS   )-1.))/2.)
  174 DENOM     =DENOM     +((N (KS   )*(N (KS   )-1.))/2.)
   46 CONTINUE
      IF(OP(4).NE.1.)GO TO 45
C  WRITE OUT THE PREFERENCE MATRIX BASED ON PROPORTIONATE DOLLAR
C  EXPENDITURE INFORMATION.
      PRINT 42,LA,(GROMAT (J)   ,J=JT,KS)
   45 CONTINUE
      IF(OP(5).NE.1.)GO TO 2
      PRINT 52
   52 FORMAT(*0   PROPORTION OF TIMES ROW LOCATION TYPES PREFERRED TO CO
     1LUMN LOCATION TYPES*/*0  -9.9 SIGNIFIES NO DATA*)
      PRINT 13
      DO 53LM=1,JZ
      JTT=((LM*(LM-1))/2)+1
      KQ=((LM*(LM-1))/2)+LM
      IF(OP(18).NE.1.)GO TO 53
C  WRITE OUT THE PREFERENCE MATRIX BASED ON WHICH OF ANY PAIR OF
C  LOCATION TYPES WAS MORE PREFERRED BY A HOUSEHOLD.
      PUNCH 180,(AGMAT(JO),JO=JTT,KQ)
  180 FORMAT(16F5.2)
   53 PRINT 42,LM,(AGMAT (JO)   ,JO=JTT,KQ)
    2 IF(OP(6).NE.1.)GO TO 3
```

```
      PRINT 190
      PRINT 13
      DO 193JG=1,JZ
      KF=((JG*(JG-1))/2)+1
      KH=((JG*(JG-1))/2)+JG
      IF(OP(19).NE.1.)GO TO 193
C   WRITE OUT A MATRIX INDICATING THE SAMPLE SIZE ASSOCIATED WITH EACH
C   PREFERENCE PROPORTION.
      PUNCH 181,(N(KU),KU=KF,KH)
  181 FORMAT(20F4.0)
  193 PRINT 192,JG,( N (KU   ),KU=KF,KH)
    3 IF(OP(15).EQ.1.)62,63
   62 IF(OPP.EQ.1.)64,63
   64 DO 60LM=1,JZ
      DO 61LU=1,LM
      LA=((LM*(LM-1))/2)+LU
      SUM(LA)=ABS(AGMAT(LA)-0.5)
      IF(AGMAT(LA).EQ.-9.9)SUM(LA)=-0.01
   61 CONTINUE
      LB=((LM*(LM-1))/2)+1
      LC=((LM*(LM-1))/2)+LM
      PUNCH 180,(SUM(LD),LD=LB,LC)
   60 CONTINUE
   63 CAGREE=((2.*GAMMA)/DENOM)-1.
C   PRINT THE GROUP≠S COEFFICIENT OF AGREEMENT.
      PRINT 55,CAGREE
   55 FORMAT(*0    COEFFICIENT OF AGREEMENT EQUALS *F5.2)
      PRINT 201
  201 FORMAT(*0    0.0 INDICATES MAXIMUM POSSIBLE DISAGREEMENT BETWEEN PR
     1EFERENCES OF HOUSEHOLDS*/*0    1.0 INDICATES TOTAL AGREEMENT BETWEE
     2N PREFERENCES OF HOUSEHOLDS*)
   13 FORMAT(*0          1   2   3   4   5   6   7   8   9  10  11  12  13
     1 14  15  16  17  18  19  20  21  22  23  24  25  26  27*)
   42 FORMAT(* *I2,2X,27F4.1)
  190 FORMAT(*0    NUMBER OF TIMES EACH LOCATION TYPE IS COMPARED TO EACH
     1 OTHER LOCATION TYPE*)
  192 FORMAT(* *I2,2X,27F4.0)
      RETURN
      END
```

```
      SUBROUTINE PEORDER(AGMAT,N,JZ,LOCTYPL,OP,OPP)
      DIMENSION AGMAT(500),N(500),TEMPL(27,27),POPNL(27,27),LOCTYPL(30),
     1SUMROWL(30),ROWSCOL(30),NROWL(30),SUM(30),DISSIM(30),OP(40)
      REAL N,NROWL
      JZZ=JZ-1
      DO 1L=1,JZ
      DO 1J=1,L
      LL=((L*(L-1))/2)+J
      TEMPL(L,J)=AGMAT(LL)
      POPNL(L,J)=N(LL)
      POPNL(J,L)=N(LL)
      TEMPL(J,L)=1.-AGMAT(LL)
      IF(AGMAT(LL).EQ.-9.9)TEMPL(J,L)=-9.9
    1 CONTINUE
      DO 258KKQ=1,JZ
  258 LOCTYPL(KKQ)=KKQ
      DO 259 JT=1,JZZ
      JG=JT+1
      DO 259JU=JG,JZ
      WW=0.
      WN=0.
      WW1=0.
      WN1=0.
C  ALGORITHM FOR REARRANGING THE ROWS AND COLUMNS OF A PREFERENCE MATRIX
      DO 261JIG=1,JZ
      IF(TEMPL(JT,JIG).EQ.-9.9.OR.TEMPL(JU,JIG).EQ.-9.9.OR.POPNL(JT,JIG)
     1.EQ.0..OR.POPNL(JU,JIG).EQ.0.)261,260
  260 WW=WW+TEMPL(JT,JIG)
      WW1=WW1+1.
      WN=WN+TEMPL(JU,JIG)
      WN1=WN1+1.
  261 CONTINUE
      IF(WW.EQ.0..OR.WW1.EQ.0.)601,602
  601 AVJT=0.
      GO TO 603
  602 AVJT=WW/WW1
  603 IF(WN.EQ.0..OR.WN1.EQ.0.)604,605
  604 AVJU=0.
      GO TO 606
  605 AVJU=WN/WN1
  606 IF(AVJT.LT.AVJU)262,259
  262 TEMPO1=LOCTYPL(JT)
      LOCTYPL(JT)=LOCTYPL(JU)
      LOCTYPL(JU)=TEMPO1
      DO 263JS=1,JZ
      TTEMPL     =TEMPL(JT,JS)
      TPOPNL     =POPNL(JT,JS)
      TEMPL(JT,JS)=TEMPL(JU,JS)
      POPNL(JT,JS)=POPNL(JU,JS)
      TEMPL(JU,JS)=TTEMPL
      POPNL(JU,JS)=TPOPNL
  263 CONTINUE
      DO 264JV=1,JZ
      TTEMPL     =TEMPL(JV,JT)
      TPOPNL     =POPNL(JV,JT)
      TEMPL(JV,JT)=TEMPL(JV,JU)
      POPNL(JV,JT)=POPNL(JV,JU)
```

```
      TEMPL(JV,JU)=TTEMPL
      POPNL(JV,JU)=TPOPNL
  264 CONTINUE
  259 CONTINUE
      IF(OP(11).NE.1.)GO TO 11
C  WRITE OUT THE REARRANGED PREFERENCE MATRIX.
      PRINT 265
  265 FORMAT(*0   PREFERENCES FOR ROW AGAINST COLUMN LOCATION TYPES WITH
     1 MOST TO LEAST PREFERRED LOCATION TYPES ORDERED FROM TOP TO BOTTOM
     2*/*0    AND LEFT TO RIGHT*)
      PRINT 119,(LOCTYPL(J),J=1,JZ)
      DO 266 JA=1,JZ
  266 PRINT 121,LOCTYPL(JA),(TEMPL(JA,JC),JC=1,JZ)
   11 IF(OP(12).NE.1.)GO TO 131
C  WRITE OUT THE REARRANGED DISSIMILARITY MATRIX.
      PRINT 106
  106 FORMAT(*0   PROXIMITY MATRIX WITH MOST TO LEAST PREFERRED LOCATION
     1 TYPES ORDERED FROM TOP TO BOTTOM AND LEFT TO RIGHT*)
      PRINT 107
  107 FORMAT(*0   -0.01 INDICATES MISSING DATA*)
      PRINT 119,(LOCTYPL(J),J=1,JZ)
  131 IF(OP(12).NE.1..AND.OP(21).NE.1.)GO TO 12
      DO 100 JA=1,JZ
      DO 110JB=1,JA
      IF(TEMPL(JA,JB).EQ.-9.9)101,102
  101 DISSIM(JB)=-0.01
      GO TO 110
  102 DISSIM(JB)=ABS(TEMPL(JA,JB)-.5)
  110 CONTINUE
      IF(OP(21).NE.1.)GO TO 132
      IF(OPP.NE.1.)GO TO 132
      PUNCH 105,LOCTYPL(JA),(DISSIM(JC),JC=1,JA)
  105 FORMAT(I2,3X,(15F5.2))
  132 IF(OP(12).NE.1.)GO TO 100
      PRINT 104,LOCTYPL(JA),(DISSIM(JC),JC=1,JA)
  104 FORMAT(1H ,I2,2X,27F4.1)
  100 CONTINUE
   12 DO 267JA=1,JZ
  267 SUM(JA)=0.
      SUMSQ=0.
      TRANS=0.
C  TRANSFORM THE PREFERENCE MATRIX TO A (1,0) MATRIX ACCORDING TO THE
C  FOLLOWING RULES. IF P IS GREATER THAN .5, IT BECOMES 1 AND IF LESS
C  THAN .5 IT BECOMES 0. IF P=.5, IT BECOMES 1 ABOVE THE MATRIX
C  DIAGONAL AND 0 BELOW IT. ALL VALUES ON THE DIAGONAL ARE SET EQUAL TO
C  0.
      DO 268 JA=1,JZ
      DO 268JW=1,JZ
      IF(JA.EQ.JW)269,270
  269 TEMPL(JA,JW)=0.
      GO TO 268
  270 IF(TEMPL(JA,JW).EQ.-9.9.OR.TEMPL(JA,JW).EQ..50)271,272
  271 IF(JA.GT.JW)273,274
  273 TEMPL(JA,JW)=0.
      GO TO 268
  274 TEMPL(JA,JW)=1.
      GO TO 275
  272 IF(TEMPL(JA,JW).GT.0.50)276,277
```

```
  277 TEMPL(JA,JW)=0.
      GO TO 268
  276 TEMPL(JA,JW)=1.
  275 SUM(JA)=SUM(JA)+TEMPL(JA,JW)
  268 CONTINUE
C     REARRANGE THE (1,0) MATRIX TO OBTAIN AS CLOSE A FIT TO PERFECT
C     TRIANGULARITY AS POSSIBLE.
      DO 50J=1,JZZ
      JG=J+1
      DO 50JC=JG,JZ
      IF(SUM(JC).GT.SUM(J))51,50
   51 TEMPO1=LOCTYPL(J)
      TEMPO2=SUM(J)
      LOCTYPL(J)=LOCTYPL(JC)
      SUM(J)=SUM(JC)
      LOCTYPL(JC)=TEMPO1
      SUM(JC)=TEMPO2
      DO 52JS=1,JZ
      TTEMPL=TEMPL(J,JS)
      TEMPL(J,JS)=TEMPL(JC,JS)
      TEMPL(JC,JS)=TTEMPL
   52 CONTINUE
      DO 53JT=1,JZ
      TTEMPL=TEMPL(JT,J)
      TEMPL(JT,J)=TEMPL(JT,JC)
      TEMPL(JT,JC)=TTEMPL
   53 CONTINUE
   50 CONTINUE
C     WRITE OUT THE (1,0) MATRIX.
      DO 278JA=1,JZ
  278 SUMSQ=SUMSQ+(SUM(JA)**2.)
      TRANS=(JZ/12.)*(JZ-1)*((2*JZ)-1)-(SUMSQ/2.)
      AMAXINT=(JZ*((JZ**2)-1))/24.
      CTRANS=TRANS/AMAXINT
      IF(OP(13).NE.1.)GO TO 13
      PRINT 279
  279 FORMAT(*0    TRANSITIVITY MATRIX*)
      PRINT 119,(LOCTYPL(J),J=1,JZ)
      DO 280JTR=1,JZ
  280 PRINT232,LOCTYPL(JTR),(TEMPL(JTR,JUR),JUR=1,JZ),SUM(JTR)
   13 PRINT 281,CTRANS
  281 FORMAT(*0    COEFFICIENT OF TRANSITIVITY IS*F6.3)
  119 FORMAT(1H0,4X,27I4)
  121 FORMAT(1H ,I2,2X,27F4.1)
  232 FORMAT(1H ,I2,2X,27F4.0,4X,F4.0)
      RETURN
      END
```

```
      SUBROUTINE RECOMP(MN)
      DIMENSION A(27,27),B(27,27),C(27,27),D(27,27),IHIST(10)
      COMMON     NL(500),NU(500),SUMONEL(500),SUMONEU(500),AGMATL(500),
     1AGMATU(500),Z1(500),TYPTEST(500),DIFGRO(27,27),FDIFF(500),
     2NCOUNT(27,25),LB,JROW,KCOL,LVAR,CUMPROB(30),IY,JZ,JMN,PROXL(500),
     3PROXU(500),OP(40),OPPUN(4,30),NDIS(60,3)
      REAL NL,NU,NCOUNT
      PRINT 47
      PRINT 466
  466 FORMAT(*0    CALCULATED USING DIRECT AND/OR INDIRECT ESTIMATES OF T
     1HE PROPORTION*)
      PRINT 200
      DO 30L=1,JZ
      DO 30J=1,L
      LL=((L*(L-1))/2)+J
      A(L,J)=SUMONEL(LL)
      A(J,L)=NL(LL)-SUMONEL(LL)
      B(L,J)=NL(LL)
      B(J,L)=NL(LL)
      IF(LVAR.EQ.0)GO TO 30
      C(L,J)=SUMONEU(LL)
      C(J,L)=NU(LL)-SUMONEU(LL)
      D(L,J)=NU(LL)
      D(J,L)=NU(LL)
   30 CONTINUE
      DO 1 LL=1,500
      SUMONEL(LL)=0.
      SUMONEU(LL)=0.
      NL(LL)=0.
    1 NU(LL)=0.
C   TO ESTIMATE P(LJ) USING AN ADITIVITY RULE ON P(LI) AND P(IJ).
C   TEST IF INFORMATION ON P(LI) AND P(IJ) EXISTS FOR THE TOTAL SAMPLE
C   OR A GROUP.
      DO 60L=1,JZ
      DO 10J=1,L
      LL=((L*(L-1))/2)+J
      IF(L.EQ.J)GO TO 10
   37 DO 40 I=1,JZ
      IF(I.EQ.L)GO TO 40
      IF(J.EQ.I)13,14
   13 SUMONEL(LL)=SUMONEL(LL)+A(L,J)
      NL(LL)=NL(LL)+B(L,J)
      IF(LVAR.EQ.0)GO TO 40
      SUMONEU(LL)=SUMONEU(LL)+C(L,J)
      NU(LL)=NU(LL)+D(L,J)
      GO TO 40
C   TEST IF THE ADDITIVITY RULE≠S ASSUMPTIONS ARE MET BY THE VALUES OF
C   P(LI) AND P(IJ).
   14 IF(B(L,I).EQ.0..OR.B(I,J).EQ.0.)GO TO 40
      IF(LVAR.EQ.0)GO TO 70
      IF(D(L,I).EQ.0..OR.D(I,J).EQ.0.)GO TO 40
   70                                  IF((A(L,I)/B(L,I)).LT..5.AND.(A(I,
     1J)/B(I,J)).EQ.1..OR.(A(L,I)/B(L,I)).EQ.1..AND.(A(I,J)/B(I,J)).LT..
     25.OR.(A(L,I)/B(L,I)).GT..5.AND.(A(I,J)/B(I,J)).EQ.0..OR.(A(L,I)/B(
     3L,I)).EQ.0..AND.(A(I,J)/B(I,J)).GT..5)GO TO 40
      IF(LVAR.EQ.0)GO TO 71
                                       IF((C(L,I)/D(L,I)).LT..5.AND.(C(I,
```

```
     1J)/D(I,J)).EQ.1..OR.(C(L,I)/D(L,I)).EQ.1..AND.(C(I,J)/D(I,J)).LT..
     25.OR.(C(L,I)/D(L,I)).GT..5.AND.(C(I,J)/D(I,J)).EQ.0..OR.(C(L,I)/D(
     3L,I)).EQ.0..AND.(C(I,J)/D(I,J)).GT..5)GO TO 40
  71 IF(B(L,I).LT.B(I,J))24,25
  24 BMIN=B(L,I)
     GO TO 26
  25 BMIN=B(I,J)
  26 IF(LVAR.EQ.0)GO TO 29
     IF(D(L,I).LT.D(I,J))27,28
  27 DMIN=D(L,I)
     GO TO 29
  28 DMIN=D(I,J)
  29 IF((A(L,I)/B(L,I)).EQ.1..AND.(A(I,J)/B(I,J)).GE..5.OR.(A(L,I)/B(L,
     1I)).GE..5.AND.(A(I,J)/B(I,J)).EQ.1.)15,17
  15 SUMONEL(LL)=SUMONEL(LL)+BMIN
     NL(LL)=NL(LL)+BMIN
     GO TO 61
  17 IF((A(L,I)/B(L,I)).EQ.0..AND.(A(I,J)/B(I,J)).LE..5.OR.(A(L,I)/B(L,
     1I)).LE..5.AND.(A(I,J)/B(I,J)).EQ.0.)18,16
  18 NL(LL)=NL(LL)+BMIN
     GO TO 61
C    LUCE≠S ADDITIVITY RULE FOR COMPUTING P(LJ)(RLJ) FROM P(LI)(A(LI
C    LUCE≠S ADDITIVITY RULE FOR COMPUTING P(LJ) I.E. RLJ, FROM P(LI) I.E.
C    A(L,I)/B(L,I), AND P(IJ) I.E. A(I,J)/B(I,J).  APPLIES TO LOWER GROUP
C    OR TOTAL SAMPLE.
  16 RLJ=((A(L,I)/B(L,I))/(1.-(A(L,I)/B(L,I))))*((A(I,J)/B(I,J))/(1.-(A
     1(I,J)/B(I,J))))
     SUMONEL(LL)=SUMONEL(LL)+(BMIN*(RLJ/(1.+RLJ)))
     NL(LL)=NL(LL)+BMIN
  61 IF(LVAR.EQ.0)GO TO 40
     IF((C(L,I)/D(L,I)).EQ.1..AND.(C(I,J)/D(I,J)).GE..5.OR.(C(L,I)/D(L,
     1I)).GE..5.AND.(C(I,J)/D(I,J)).EQ.1.)62,63
  62 SUMONEU(LL)=SUMONEU(LL)+DMIN
     NU(LL)=NU(LL)+DMIN
     GO TO 40
  63 IF((C(L,I)/D(L,I)).EQ.0..AND.(C(I,J)/D(I,J)).LE..5.OR.(C(L,I)/D(L,
     1I)).LE..5.AND.(C(I,J)/D(I,J)).EQ.0.)64,65
  64 NU(LL)=NU(LL)+DMIN
     GO TO 40
C    LUCE≠S ADDITIVITY RULE(UPPER GROUP).
  65 RLJ=((C(L,I)/D(L,I))/(1.-(C(L,I)/D(L,I))))*((C(I,J)/D(I,J))/(1.-(C
     1(I,J)/D(I,J))))
     SUMONEU(LL)=SUMONEU(LL)+(DMIN*(RLJ/(1.+RLJ)))
     NU(LL)=NU(LL)+DMIN
  40 CONTINUE
  10 CONTINUE
  60 CONTINUE
C    OBTAIN RECOMPUTED P(LJ)(FOR EACH CELL OF MATRIX), USING THE AVERAGE
C    OF ALL THE INDIRECT ESTIMATES OF P(LJ) CALCULATED ABOVE.
     DO 2L=1,JZ
     DO 3J=1,L
     LL=((L*(L-1))/2)+J
     IF(J.EQ.L)50,51
  50 AGMATL(LL)=.5
     AGMATU(LL)=.5
     GO TO 3
  51 IF(SUMONEL(LL).EQ.0..AND.NL(LL).EQ.0.)4,5
   4 AGMATL(LL)=-9.9
```

```
      GO TO 72
    5 AGMATL(LL)=SUMONEL(LL)/NL(LL)
   72 IF(LVAR.EQ.0)GO TO 3
      IF(SUMONEU(LL).EQ.0..AND.NU(LL).EQ.0.)73,74
   73 AGMATU(LL)=-9.9
      GO TO 3
   74 AGMATU(LL)=SUMONEU(LL)/NU(LL)
    3 CONTINUE
      LA=((L*(L-1))/2)+1
      LB=((L*(L-1))/2)+L
      IF(OP(7).NE.1.)GO TO 180
C     WRITE OUT RECOMPUTED PREFERENCE MATRICES FOR BOTH GROUPS OR TOTAL
C     SAMPLE.
      PRINT 6,L,(AGMATL(LU),LU=LA,LB)
    6 FORMAT(* *I2,2X,27F4.1)
      IF(LVAR.EQ.0)GO TO 180
      PRINT 6,L,(AGMATU(LU),LU=LA,LB)
  180 IF(OP(20).NE.1..OR.OPPUN(1,MN).NE.1.)GO TO 2
      PUNCH 555,(AGMATL(LU),LU=LA,LB)
    2 CONTINUE
      IF(LVAR.NE.C.AND.OP(20).EQ.1..AND.OPPUN(1,MN).EQ.1.)181,182
  181 DO 183 L=1,JZ
      LA=((L*(L-1))/2)+1
      LB=((L*(L-1))/2)+L
  183 PUNCH 555,(AGMATU(LU),LU=LA,LB)
  182 IF(OP(8).NE.1.)GO TO 79
      PRINT 190
  190 FORMAT(*0    NUMBER OF TIMES EACH LOCATION TYPE IS COMPARED TO EACH
     1 OTHER LOCATION TYPE*)
      PRINT 466
      PRINT 200
  200 FORMAT(*0          1    2    3    4    5    6    7    8    9   10   11   12   13
     1 14   15   16   17   18   19   20   21   22   23   24   25   26   27*)
C     WRITE OUT THE SAMPLE SIZE  ASSOCIATED WITH EACH RECOMPUTED P(LJ) FOR
C     COMPUTE AND WRITE OUT PROXIMITY MATRICES BASED ON THE NEW PREFERENCE
C     MATRICES.
      DO 7L=1,JZ
      LA=((L*(L-1))/2)+1
      LB=((L*(L-1))/2)+L
      PRINT 192,L,(NL(LU),LU=LA,LB)
      IF(LVAR.EQ.0)GO TO 79
      PRINT 192,L,(NU(LU),LU=LA,LB)
    7 CONTINUE
   79 IF(OP(9).NE.1.)GO TO 80
      PRINT 554
  554 FORMAT(*0    PROXIMITY MATRIX*/*  -9.9 SIGNIFIES NO DATA*)
      PRINT 200
   80 DO 552  L=1,JZ
      DO 553 JJH=1,L
      LL=((L*(L-1))/2)+JJH
      PROXL(LL)=ABS(AGMATL(LL)-.5)
      IF(AGMATL(LL).EQ.-9.9)PROXL(LL)=-9.9
      IF(LVAR.EQ.0)GO TO 553
      PROXU(LL)=ABS(AGMATU(LL)-.5)
      IF(AGMATU(LL).EQ.-9.9)PROXU(LL)=-9.9
  553 CONTINUE
      JA=((L*(L-1))/2)+1
      JB=((L*(L-1))/2)+L
```

```
      IF(OP(9).NE.1.)GO TO 81
      PRINT   6,L,(PROXL(LU),LU=JA,JB)
   81 M=L
      IF(OP(10).NE.1.)GO TO 552
      IF(OPPUN(1,MN).NE.1.)GO TO 552
      DO 556 JJI=1,M
      LL=((M*(M-1))/2)+JJI
      IF(LVAR.EQ.0)GO TO 556
      IF(PROXU(LL).EQ.-9.9)PROXU(LL)=-0.01
  556 IF(PROXL(LL).EQ.-9.9)PROXL(LL)=-0.01
      PUNCH 555,(PROXL(LU),LU=JA,JB)
  555 FORMAT(16F5.2)
  552 CONTINUE
      IF(LVAR.EQ.0)GO TO 82
      DO 83L=1,JZ
      JA=((L*(L-1))/2)+1
      JB=((L*(L-1))/2)+L
      IF(OP(9).NE.1.)GO TO 84
      PRINT   6,L,(PROXU(LU),LU=JA,JB)
   84 IF(OP(10).NE.1.)GO TO 83
      IF(OPPUN(1,MN).NE.1.)GO TO 83
      PUNCH 555,(PROXU(LU),LU=JA,JB)
   83 CONTINUE
   82 IF(LVAR.EQ.0)210,211
C     COMPUTE A HISTOGPAM OF PROXIMITY VALUES FOR ONE OR BOTH GROUPS.
  210 JS=1
      GO TO 213
  211 JS=2
  213 DO 75LLL=1,JS
      DO 34L=1,10
   34 IHIST(L)=0
   76 DO 78I=1,JMN
      IF(LLL.NE.2)GO TO 77
      PROXL(I)=PROXU(I)
   78 CONTINUE
   77 DO 31J=1,JMN
      IF(PROXL(J).LT.0.)GO TO 31
      R=0.0
      DO 32L=1,9
      R=R+0.05
      IF(PROXL(J).LE.R)33,32
   33 IHIST(L)=IHIST(L)+1
      GO TO 31
   32 CONTINUE
      IHIST(10)=IHIST(10)+1
   31 CONTINUE
  214 PRINT 36
   36 FORMAT(*0     FREQUENCY DISTRIBUTION OF PROXIMITY VALUES*/*0     .00
     1     .05     .10     .15     .20     .25     .30     .35     .40     .45
     2  .50*)
   75 PRINT 35,(IHIST(J),J=1,10)
   35 FORMAT(1H0,4X,10(3X,I4))
  192 FORMAT(* *I2,2X,27F4.0)
   47 FORMAT(*0   PROPORTION OF TIMES ROW LOCATION TYPE PREFERRED TO COL
     1UMN LOCATION TYPE*/*0    -9.9 SIGNIFIES NO DATA*)
      RETURN
      END
```

```
      SUBROUTINE CHITEST
      COMMON      NL(500),NU(500),SUMONEL(500),SUMONEU(500),AGMATL(500),
     1AGMATU(500),Z1(500),TYPTEST(500),DIFGRO(27,27),FDIFF(500),
     2NCOUNT(27,25),LB,JPOW,KCOL,LVAR,CUMPROB(30),IY,JZ,JMN,PROXL(500),
     3PROXU(500),OP(40),OPPUN(4,30),NDIS(60,3)
      DIMENSION A(4)
      INTEGER A
      REAL NU,NL,L1,L2,NCOUNT
      IF(AGMATL(LB).EQ.AGMATU(LB))GO TO 41
C     ASSIGN INTEGER VALUES TO A TWO BY TWO CONTINGENCY TABLE, SHOWING THE
C     NUMBER IN EACH GROUP ACCEPTING EACH OF THE PAIR OF LOCATION TYPES(X,
C     XX,Y,YY).
      IF(SUMONEL(LB).EQ.0.)GO TO 55
      J=SUMONEL(LB)
      S=SUMONEL(LB)-J
      IF(S.EQ.0.)55,74
   74 IF(SUMONEL(LB).LT.(NL(LB)/2.))54,55
   54 A(2)=SUMONEL(LB)+1.
      GO TO 56
   55 A(2)=SUMONEL(LB)
   56 A(1)=NL(LB)-A(2)
      IF(SUMONEU(LB).EQ.0.)GO TO 58
      J=SUMONEU(LB)
      S=SUMONEU(LB)-J
      IF(S.EQ.0.)58,75
   75 IF(SUMONEU(LB).LT.(NU(LB)/2.))57,58
   57 A(4)=SUMONEU(LB)+1.
      GO TO 59
   58 A(4)=SUMONEU(LB)
   59 A(3)=NU(LB)-A(4)
      X=A(1)
      XX=A(2)
      Y=A(3)
      YY=A(4)
      IF(X.EQ.0..AND.Y.EQ.0..OR.XX.EQ.0..AND.YY.EQ.0.)61,71
   71 IF(X.EQ.0..OR.XX.EQ.0..OR.Y.EQ.0..OR.YY.EQ.0.)62,72
   72 IF((X/XX).EQ.(Y/YY))61,62
   61 Z1(LB)=0.
      TYPTEST(LB)=4RCHI2
       FDIFF(LB)=0.
      GO TO 44
   62 R1=X/NL(LB)
      R2=Y/NU(LB)
      IF(R1.GT.R2)5,6
    5 KU=1
      GO TO 63
    6 KU=2
   63 DO 10J=1,100
C     COMPUTE CHI SQUARED ACCORDING TO THE FORMULA GIVEN BY SIEGEL,S. IN
C     NONPARAMETRIC STATISTICS FOR THE BEHAVIORAL SCIENCES,PAGE 107.
      Z1(LB)     =((NL(LB)+NU(LB))*(ABS(( X*YY)-( XX*Y))-((NL(LB)+NU(LB))
     1/2.))**2)/(( X+XX)*( Y+YY)*( X+Y )*(XX+YY))
      IF(J.EQ.1)42,43
   42 CALL DSTRB(0)
C     TEST IF THE COMPUTED CHI SQUARED VALUE EXCEEDS THE CONFIDENCE LIMIT
C     OF 3.84 IN A ONE-TAILED TEST WITH ALPHA=.05. IF IT EXCEEDS THAT
C     VALUE, REPEAT THE TEST WITH PROGRESSIVELY LESS EXTREME DISTRIBUTIONS
```

```
C   OF VALUES IN THE TWO BY TWO TABLE, WHILST KEEPING MARGINAL TOTALS
C   FIXED, UNTIL THAT DISTRIBUTION IS OBTAINED WHICH RESULTS IN CHI
C   SQUARED BEING LESS THAN OR EQUAL TO 3.84.
    43 IF(Z1(LB).GT.3.84)1,2
     1 TEMP=Z1(LB)
       TEMP1=ABS(X/NL(LB)-Y/NU(LB))
       GO TO (8,9),KU
     8 X=X-1.
       XX=XX+1.
       Y=Y+1.
       YY=YY-1.
       GO TO 21
     9 X=X+1.
       XX=XX-1.
       Y=Y-1.
       YY=YY+1.
    21 IF((R1-R2)*(X/NL(LB)-Y/NU(LB)).LE.0.)4,10
     2 IF(J.EQ.1)3,4
     3 TYPTEST(LB)=4RCHI2
       FDIFF(LB)=0.
       RETURN
     4 Z1(LB)=TEMP
       TYPTEST(LB)=5R*CHI2
C   COMPUTE THE SIZE OF THE GAP BETWEEN THE ACTUAL DIFFERENCE BETWEEN
C   THE TWO GROUPS≠ P ESTIMATES AND THAT HYPOTHETICAL DIFFERENCE(GIVEN
C   THE SAME MARGINAL TOTALS ) WHICH WAS JUST SIGNIFICANTLY DIFFERENT.
       FDIFF(LB)=ABS(ABS(DIFGRO(JROW,KCOL))-TEMP1)
       NCOUNT(KCOL,LVAR)=NCOUNT(KCOL,LVAR)+1.
       NCOUNT(JROW,LVAR)=NCOUNT(JROW,LVAR)+1.
       RETURN
    10 CONTINUE
       RETURN
    41 Z1(LB)=0.
       TYPTEST(LB)=4RCHI2
    44 CALL DSTRB(0)
       RETURN
       END
```

```
      SUBROUTINE FEPTEST
      COMMON      NL(500),NU(500),SUMONEL(500),SUMONEU(500),AGMATL(500),
     1AGMATU(500),Z1(500),TYPTEST(500),DIFGRO(27,27),FDIFF(500),
     2NCOUNT(27,25),LB,JPOW,KCOL,LVAP,CUMPROB(30),IY,JZ,JMN,PROXL(500),
     3PROXU(500),OP(40),OPPUN(4,30),NDIS(60,3)
      DIMENSION A(4),GG(5),RCSUM(4),NUM(4),DENOM(5),I(4)
      REAL NU,NL   ,NUM,NCOUNT
      INTEGER A,GG,RCSUM,TEMPZ,AMIN
      IF(AGMATL(LB).EQ.AGMATU(LB))GO TO 103
C     ASSIGN INTEGER VALUES TO THE TWO BY TWO CONTINGENCY TABLE, SHOWING
C     THE NUMBER IN EACH GROUP ACCEPTING EACH LOCATION TYPE(A(1) TO A(4)).
      IF(SUMONEL(LB).EQ.0.)GO TO 55
      J=SUMONEL(LB)
      S=SUMONEL(LB)-J
      IF(S.EQ.0.)55,74
   74 IF(SUMONEL(LB).LT.(NL(LB)/2.))54,55
   54 A(2)=SUMONEL(LB)+1.
      GO TO 56
   55 A(2)=SUMONEL(LB)
   56 A(1)=NL(LB)-A(2)
      IF(SUMONEU(LB).EQ.0.)GO TO 58
      J=SUMONEU(LB)
      S=SUMONEU(LB)-J
      IF(S.EQ.0.)58,75
   75 IF(SUMONEU(LB).LT.(NU(LB)/2.))57,58
   57 A(4)=SUMONEU(LB)+1.
      GO TO 59
   58 A(4)=SUMONEU(LB)
   59 A(3)=NU(LB)-A(4)
      X=A(1)
      XX=A(2)
      Y=A(3)
      YY=A(4)
      IF(X.EQ.0..AND.Y.EQ.0..OR.XX.EQ.0..AND.YY.EQ.0.)61,71
   71 IF(X.EQ.0..OR.XX.EQ.0..OR.Y.EQ.0..OR.YY.EQ.0.)62,72
   72 IF((X/XX).EQ.(Y/YY))61,62
   61 Z1(LB)=1.
      TYPTEST(LB)=3RFEP
      GO TO 104
   62 R1=X/NL(LB)
      R2=Y/NU(LB)
      IF(R1.GT.R2)42,43
   42 KW=1
      GO TO 63
   43 KW=2
C     COMPUTE INITIAL VALUES IN THE FORMULA FOR COMPUTING FISHER≠S EXACT
C     PROBABILITY. SEE SIEGEL,S.,NONPARAMETRIC STATISTICS FOR THE
C     BEHAVIORAL SCIENCES, PAGE 97, EQUATION 6.1.
   63 GG(1)=NL(LB)+NU(LB)
      ALPHA=0.05
      RCSUM(1)=A(1)+A(2)
      RCSUM(2)=A(3)+A(4)
      RCSUM(3)=A(1)+A(3)
      RCSUM(4)=A(2)+A(4)
      DO 9J=1,3
      L=J+1
      DO 9K=L,4
```

```
      IF(RCSUM(K).GT.RCSUM(J))11,9
   11 TEMPZ=RCSUM(J)
      RCSUM(J)=RCSUM(K)
      RCSUM(K)=TEMPZ
    9 CONTINUE
      DO 100JI=1,100
      DO 60J=1,4
   60 I(J)=A(J)
C     REARRANGE VALUES IN THE NUMERATOR AND DENOMINATOR IN ORDER OF
C     MAGNITUDE TO MAXIMISE CANCELLATION BETWEEN NUMERATOR AND DENOMINATOR
C     AND THEREFORE MINIMISE THE SUBSEQUENT CALCULATION OF FACTORIALS.
      GG(2)=A(1)
      GG(3)=A(2)
      GG(4)=A(3)
      GG(5)=A(4)
      AMIN=A(1)
      DO 27K=2,4
      IF(A(K).LT.AMIN)AMIN=A(K)
   27 CONTINUE
   50 JMIN=AMIN+1
      P1=0.
      PP=0.
      KU=1
      DO 10J=1,JMIN
      IF(JMIN.EQ.1)KU=4
      DO 14JJ=1,4
      NUM(JJ)=1.
   14 DENOM(JJ)=1.
      DENOM(5)=1.
      DO 26JL=2,4
      LL=JL+1
      DO 26L=LL,5
      IF(GG(L).GT.GG(JL))28,26
   28 TEMPZ=GG(JL)
      GG(JL)=GG(L)
      GG(L)=TEMPZ
   26 CONTINUE
C     CANCEL OUT BETWEEN NUMERATOR AND DENOMINATOR AND COMPUTE THE
C     FACTORIALS OF THE VALUES REMAINING IN EACH.
      DO 15JJ=1,4
      IF(RCSUM(JJ).GT.GG(JJ))16,17
   16 KK=RCSUM(JJ)-GG(JJ)
      DO 20JV=1,KK
      NUM(JJ)=NUM(JJ)*RCSUM(JJ)
      RCSUM(JJ)=RCSUM(JJ)-1
   20 CONTINUE
      GO TO 15
   17 IF(RCSUM(JJ).EQ.GG(JJ))15,18
   18 KK=GG(JJ)-RCSUM(JJ)
      DO 21JV=1,KK
      DENOM(JJ)=DENOM(JJ)*GG(JJ)
      GG(JJ)=GG(JJ)-1
   21 CONTINUE
   15 CONTINUE
      KK=GG(5)-1
      IF(KK.LE.0)34,35
   35 DO 37LU=1,KK
      DENOM(5)=DENOM(5)*GG(5)
```

```
      GG(5)=GG(5)-1
   37 CONTINUE
C     COMPUTE THE PRODUCT OF THE VALUES IN THE NUMERATOR AND DENOMINATOR
C     OF THE FORMULA.
   34 PP=PP+(NUM(1)*NUM(2)*NUM(3)*NUM(4))/(DENOM(1)*DENOM(2)*DENOM(3)*DE
     1NOM(4)*DENOM(5))
      IF(J.EQ.1)22,23
   22 IF(JMIN.EQ.1)101,102
  102 P1=PP
      PP=0.
  101 GO TO(23,1,2,10),KU
C     CHECK WHICH CELL VALUES SHOULD BE INCREMENTED  AND DECREMENTED BY 1
C     SO THAT THE MOST EXTREME DISTRIBUTION OF CELL VALUES, WHERE ONE VALUE
C     EQUALS ZERO, WILL BE ACHIEVED IN AS FEW STEPS AS POSSIBLE.
   23 IF(I(1).EQ.AMIN.AND.I(2).EQ.AMIN)76,77
   76 IF(I(3).GT.I(4))1,2
   77 IF(I(3).EQ.AMIN.AND.I(4).EQ.AMIN)78,79
   78 IF(I(1).GT.I(2))2,1
   79 IF(I(1).EQ.AMIN.AND.I(4).EQ.AMIN)1,80
   80 IF(I(2).EQ.AMIN.AND.I(3).EQ.AMIN)2,81
   81 IF(I(1).EQ.AMIN.AND.I(3).EQ.AMIN)82,83
   82 IF(I(2).GT.I(4))1,2
   83 IF(I(2).EQ.AMIN.AND.I(4).EQ.AMIN)84,85
   84 IF(I(1).LT.I(3))1,2
   85 IF(I(1).EQ.AMIN.OR.I(4).EQ.AMIN)1,2
C     DECREMENT AND INCREMENT THE APPROPRIATE CELL VALUES.
    1 KU=2
      A(1)=A(1)-1.
      A(2)=A(2)+1.
      A(3)=A(3)+1.
      A(4)=A(4)-1.
      GO TO 40
    2 KU=3
      A(1)=A(1)+1.
      A(2)=A(2)-1.
      A(3)=A(3)-1.
      A(4)=A(4)+1.
   40 GG(2)=A(1)
      GG(3)=A(2)
      GG(4)=A(3)
      GG(5)=A(4)
C     RETURN TO BEGINNING OF LOOP AND COMPUTE THE PROBABILITY ASSOCIATED
C     WITH EACH MORE EXTREME DISTRIBUTION.
   10 CONTINUE
C     CHECK IF THE CUMULATIVE PROBABILITY COMPUTED EXCEEDS ALPHA(.05) YET.
      IF(PP.GT.ALPHA)91,92
   91 IF(JI.GT.1)96,97
   97 Z1(LB)=PP+P1
      TYPTEST(LB)=3PFEP
   98 FDIFF(LB)=0.
      GO TO 104
C     IF THE SUM OF ALL PROBABILITIES EXCEPT THE LARGEST IS LESS THAN ALPHA
C     , BUT EXCEEDS ALPHA IF THE LARGEST IS INCLUDED, USE TOCHER≠S
C     MODIFICATION TO DECIDE WHETHER TO ACCEPT THE PROPORTIONS AS
C     SIGNIFICANTLY DIFFERENT OR NOT.
   92 IF((P1+PP).LE.ALPHA)94,93
   93 R=(ALPHA-PP)/P1
      CALL RANDU(IY,IY,T)
```

```
      IF(T.LT.R)95,91
   95 Z1(LB)=PP+P1
      TYPTEST(LB)=4R*FEP
      NCOUNT(JROW,LVAR)=NCOUNT(JROW,LVAR)+1.
      NCOUNT(KCOL,LVAR)=NCOUNT(KCOL,LVAR)+1.
      IF(JT.EQ.1)98,99
C     AS IN THE CHI SQUARED TEST, IF THE ACTUAL PROPORTIONS ARE
C     SIGNIFICANTLY DIFFERENT, REPEAT THE TEST WITH PROGRESSIVELY LESS
C     EXTREME HYPOTHETICAL DISTRIBUTIONS OF VALUES  IN THE TWO BY TWO
C     TABLE(WHILST KEEPING THE MARGINAL TOTALS FIXED), UNTIL THAT
C     DISTRIBUTION IS OBTAINED WHICH RESULTS IN FISHER≠S EXACT PROBABILITY
C     EXCEEDING .05.
   94 TEMP=PP+P1
      TEMP1=ABS(X/NL(LB)-Y/NU(LB))
      GO TO (44,45),KW
   44 A(1)=A(1)-1.
      A(2)=A(2)+1.
      A(3)=A(3)+1.
      A(4)=A(4)-1.
      GO TO 46
   45 A(1)=A(1)+1.
      A(2)=A(2)-1.
      A(3)=A(3)-1.
      A(4)=A(4)+1.
   46 IF((R1-R2)*(A(1)/NL(LB)-A(3)/NU(LB)).LE.0.)96,100
   96 Z1(LB)=TEMP
      TYPTEST(LB)=4R*FEP
      NCOUNT(JROW,LVAR)=NCOUNT(JROW,LVAR)+1.
      NCOUNT(KCOL,LVAR)=NCOUNT(KCOL,LVAR)+1.
C     AS IN THE CHI SQUARED TEST COMPUTE THE SIZE OF THE GAP BETWEEN THE
C     ACTUAL DIFFERENCEBETWEEN THE TWO GROUP≠S P ESTIMATES AND THAT
C     HYPOTHETICAL DIFFERENCE(GIVEN THE SAME MARGINAL TOTALS), WHICH WAS
C     JUST SIGNIFICANT.
   99 FDIFF(LB)=ABS(ABS(DIFGRO(JROW,KCOL))-TEMP1)
      CALL DSTRB(1)
      RETURN
  100 CONTINUE
      RETURN
  103 Z1(LB)=1.
      TYPTEST(LB)=3RFEP
  104 CALL DSTRB(1)
      RETURN
      END
```

```
      SUBROUTINE DSTRB(N3)
      COMMON      NL(500),NU(500),SUMONEL(500),SUMONEU(500),AGMATL(500),
     1AGMATU(500),Z1(500),TYPTEST(500),DIFGRO(27,27),FDIFF(500),
     2NCOUNT(27,25),LB,JPOW,KCOL,LVAR,CUMPROB(30),IY,JZ,JMN,PROXL(500),
     3PROXU(500),OP(40),OPPUN(4,30),NDIS(60,3)
C   CALCULATE THE DISTRIBUTION OF STATISTICAL VALUES FOUND IN EACH TYPE
C   OF STATISTICAL TEST.
      IF(N3)2,2,3
C   CALCULATE A HISTOGRAM OF CHI SQUARED VALUES BETWEEN 0 AND 10 IN
C   INCREMENTS OF .2.
    2 X=0.
      Y=.2
      K=2
      N=50
      GO TO 100
C   CALCULATE A HISTOGRAM OF EXACT PROBABILITY VALUES BETWEEN 0 AND 1 IN
C   INCREMENTS OF .05.
    3 X=0.
      Y=.05
      K=3
      N=20
  100 DO 10J=1,N
      X=X+Y
      IF(Z1(LB).LT.X)9,10
    9 NDIS(J,K)=NDIS(J,K)+1
      RETURN
   10 CONTINUE
      NP1=N+1
      NDIS(NP1,K)=NDIS(NP1,K)+1
      RETURN
      END
```

Program: HYPBEH

This program is designed to compute household's choices of towns which would be consistent with given deterministic and probabilistic spatial preference rules. Having computed the town(s) chosen by a household, it then allocates them and those towns rejected to location types in the same fashion as program LOCTYPE.

```
      PROGRAM HYPBEH(INPUT,OUTPUT,TAPE1,TAPE5=INPUT,TAPE6=OUTPUT)
      COMMON ITD(1500),IE(1500),IN(1500),IPOP(1500),
     1IIDVIS(9),LIMPOP(7),LIMDIS(7,4),          DDIST(50)
      COMMON IY
      DIMENSION IHE(600),IHN(600),IHD(600),NT(600),I1(10),I2(10)    ,
     1NRULE(20),EXP(20,5),CUMPRO(20,5),EXPM(20),EXPSD(20),DETPRO(20)
      DIMENSION I(40),II(20)
      INTEGER DETPRO,HE,HN
C   READ IN DATA DESCRIBING EACH TOWN≠S POPULATION(IPOP,EASTING(IE), AND
C   NORTHING(IN).
   28 READ 1,(ITD(J),IE(J),IN(J),IPOP(J),J=1,1342)
    1 FORMAT(5(I4,2I3,I6))
      J=1
C   READ IN A HOUSEHOLD≠S EASTING(IHE),NORTHING(IHN) AND THE NUMBER OF
C   TOWNS PATRONISED(JB).
    8 READ 2,JB,IHD(J),IHE(J),IHN(J),(I1(K),I2(K),K=1,JB)
    2 FORMAT(2X,I2,2X,3I4,(3(I5,6X,I7)))
   26 IF(IHD(J).GE.900)5,7
    7 NT(J)=JB
      J=J+1
      GO TO 8
    5 NHOUS=J-1
C   READ IN THE DISTANCE AND POPULATION LIMITS(LIMDIS,LIMPOP) OF EACH
C   POSSIBLE LOCATION TYPE.
      READ 4,(LIMPOP(J),J=1,7)
    4 FORMAT(7I8)
      PRINT 31,(LIMPOP(J),J=1,7)
   31 FORMAT(1X,7I8)
      READ 6,(LIMDIS(1,K),K=1,3)
    6 FORMAT(4I4)
      PRINT 31,(LIMDIS(1,K),K=1,3)
      READ 6,((LIMDIS(J,K),K=1,4),J=2,7)
      PRINT 31,((LIMDIS(J,K),K=1,4),J=2,7)
C   READ IN THE NUMBER OF SETS OF PREFERENCE RULES FOR WHICH BEHAVIOR
C   IS TO BE GENERATED.
      READ 21,NLOOP
      PRINT21,NLOOP
      DO 10 J=1,NLOOP
C   FOR EACH SET, SPECIFY WHETHER THE RULES ARE DETERMINISTIC OR
C   PROBABILISTIC(DETPRO) AND THE NUMBER OF PREFERENCE RULES IN THE SET
C   (NRULE).
      READ 22, DETPRO(J),NRULE(J)
      PRINT22, DETPRO(J),NRULE(J)
      N=NRULE(J)
C   READ IN EACH OF THE PREFERENCE RULES(EXP) WHERE THE BASIC RULE IS A
C   GRAVITY MODEL OF THE FORM P/D**EXP WHERE P=TOWN POPULATION, D=
C   DISTANCE FROM HOUSEHOLD TO TOWN. ALSO READ IN THE PROPORTION OF THE
C   TOTAL SAMPLE TO BE RANDOMLY ASSIGNED TO EACH OF THE NRULE RULES.
      READ 23,(EXP(J,K),CUMPRO(J,K),K=1,N)
      PRINT23,(EXP(J,K),CUMPRO(J,K),K=1,N)
   10 CONTINUE
C   READ IN THE TOTAL NUMBER OF SETS OF PREFERENCE RULES INCLUDING THE
C   NUMBER IN THE FIRST SET.
      READ 21,NTOT
      PRINT21,NTOT
      IF(NTOT.EQ.NLOOP)GO TO 99
      NP1=NLOOP+1
```

```
C   SPECIFY WHETHER THE SET OF RULES IS DETERMINISTIC OR PROBABILISTIC
C   (DETPRO),THE MEAN EXPONENT VALUE FOR ALL THE RULES(EXPM) AND THE
C   EXPONENT≠S STANDARD DEVIATION(EXPSD). THE BASIC RULE IS AGAIN A
C   GRAVITY MODEL OF THE TYPE P/D**EXP, BUT IN THIS CASE THERE IS AN
C   INFINITE POSSIBLE NUMBER OF PREFERENCE RULES NORMALLY DISTRIBUTED
C   ABOUT A MEAN PREFERENCE RULE P/D**EXPM WITH A STANDARD DEVIATION OF
C   EXPSD. HOUSEHOLDS OBEYING RULES WITH AN EXPONENT LESS THAN THE MEAN
C   RULE≠S ARE ASSIGNED TO ONE GROUP AND THE REST TO ANOTHER GROUP.
        READ 24,(DETPRO(J),EXPM(J),EXPSD(J),J=NP1,NTOT)
        PRINT24,(DETPRO(J),EXPM(J),EXPSD(J),J=NP1,NTOT)
C   INITIALISE THE UNIFORMLY AND NORMALLY DISTRIBUTED RANDOM NUMBER
C   GENERATORS.
    99  IX=9
        IIX=9
        CALL RANDU(IX,IY,Y)
        CALL GAUSS(IIX,1.,2.,EX)
        DO 20 JR=1,100
        CALL GAUSS( IIX,1.,2.,EX)
    20  CALL RANDU(IY,IY,Y)
        DO 1001 JQ=1,NTOT
        IF(JQ.EQ.8.OR.JQ.EQ.10)999,1001
   999  DO 1000 KQ=1,NHOUS
        IF(JQ.GT.NLOOP)52,51
    51  CALL RANDU(IY,IY,U)
        JT=NRULE(JQ)
        DO 53 J=1,JT
        IF(U.GT.CUMPRO(JQ,J))53,54
    54  NUM=J
        EX=EXP(JQ,J)
C   CALL THE SUBROUTINE WHICH COMPUTES WHICH TOWN(S) THE HOUSEHOLD
C   PATRONISES CONSISTENT WITH THE PREFERENCE RULE RANDOMLY ASSIGNED
C   TO THE HOUSEHOLD.
        CALL RULE(IHE(KQ),IHN(KQ),NT(KQ),NUM,DETPRO(JQ),EX)
        GO TO 1000
    53  CONTINUE
        PRINT 55
    55  FORMAT(* ERROR*)
        GO TO 51
    52  EM=EXPM(JQ)
        ESD=EXPSD(JQ)
    16  CALL GAUSS(IIX,ESD,EM,EX)
        IF(EX.LT.0..OR.EX.GT.4.)13,14
    13  PRINT 15,EX,IHE(KQ),IHN(KQ)
    15  FORMAT(1X,F6.3,2I4)
        GO TO 16
    14  IF(EX.LT.EM)11,12
    11  NUM=1
        GO TO 113
    12  NUM=2
   113  CALL RULE(IHE(KQ),IHN(KQ),NT(KQ),NUM,DETPRO(JQ),EX)
  1000  CONTINUE
        PRINT 97
    97  FORMAT(1H1)
        ENDFILE 1
  1001  CONTINUE
    21  FORMAT(1X,I2)
    22  FORMAT(1X,79I1)
    23  FORMAT(1X,15F5.2)
    24  FORMAT(1X,I1,2F5.2)
    95  STOP
        END
```

```
      SUBROUTINE RULE(HE,HN,NT,NUM,DETPRO,EXP)
      COMMON ITD(1500),IE(1500),IN(1500),IPOP(1500),
     1IIDVIS(9),LIMPOP(7),LIMDIS(7,4),         DDIST(50)
      COMMON IY
      DIMENSION DISRAT(91),IDVIS(90),U(91),DIST(90)
      REAL MAXPDE,MINDIS
      INTEGER EX,DETPRO,HE,HN
      NN=10
      NNP1=NN+1
      GO TO (20,19),DETPRO
   20 NN=NT
      IF(NN.EQ.0)RETURN
   19 DO 1 K=1,NN
      IF(EXP.EQ.0.)25,26
   25 MAXPDE=999.
      GO TO 27
   26 MAXPDE=0.
   27 KK=K-1
C     FOR EACH TOWN CALCULATE IF IT LIES WITHIN THE LOCATION TYPES DEFINED.
C     IF IT DOES, CALCULATE THE P/D**EXP SCORE OF THAT TOWN. FOR A
C     DETERMINISTIC RULE FIND THE TOWN(S) WHICH HAVE THE LARGEST SCORES,
C     FOR AS MANY TOWNS AS THE HOUSEHOLD PATRONISES. FOR A PROBABILISTIC
C     RULE FIND THE TOWNS WITH THE 10 HIGHEST SCORES AND RANDOMLY SELECT
C     AS MANY TOWNS FROM THE 10 AS THE HOUSEHOLD PATRONISES, WITH EACH
C     TOWN≠S PROBABILITY OF BEING SELECTED EQUAL TO ITS P/D**EXP SCORE
C     EXPRESSED AS A PROPORTION OF THE SUM OF THE 10 SCORES.
      DO 2J=1,1342
      IF(IPOP(J).EQ.0.OR.IPOP(J).GT.LIMPOP(7))GO TO 2
      IF(K.EQ.1)GO TO 5
      DO 4KL=1,KK
      IF(IDVIS(KL).EQ.ITD(J))GO TO 2
    4 CONTINUE
    5 LX   =IABS(HE-IE(J))+IABS(HN-IN(J))
      DO 28LD=1,7
      IF(IPOP(J).LT.LIMPOP(LD))29,28
   29 IF(LD.EQ.1)30,31
   30 DO 32LE=1,3
      IF(LX.LT.LIMDIS(LD,LE))33,32
   32 CONTINUE
      GO TO 2
   31 DO 34LF=1,4
      IF(LX.LT.LIMDIS(LD,LF))33,34
   34 CONTINUE
      GO TO 2
   28 CONTINUE
   33 BX=LX
      IF(BX.EQ.0.)BX=0.001
      IF(EXP.EQ.0.)13,14
   13 COMP=BX
      IF(COMP.LT.MAXPDE)3,2
   14 BX=BX**EXP
      D=IPOP(J)*10.
      COMP=D/BX
      IF(COMP.GT.MAXPDE)3,2
    3 MAXPDE=COMP
      IMAXD=ITD(J)
    2 CONTINUE
```

```
      IDVIS(K)=IMAXD
      DIST(K)=MAXPDE
    1 CONTINUE
      GO TO (21,22),DETPRO
   21 DO 23JU=1,NT
      DDIST(JU)=DIST(JU)
   23 IIDVIS(JU)=IDVIS(JU)
C    HAVING FOUND THE TOWNS CHOSEN ACCORDING TO A DETERMINISTIC RULE, CALL
C    THE SUBROUTINE CALC WHICH ASSIGNS THESE TOWNS AT THESE DISTANCES
C    FROM THE HOUSEHOLDS TO LOCATION TYPES AND COMPUTES THE LOCATION TYPES
C    TO WHICH REJECTED TOWNS BELONG, AND THE NUMBER OF SUCH TOWNS IN EACH
C    TYPE.
      CALL CALC(NUM,NT,HE,HN)
      RETURN
   22 U(1)=0.
      IF(EXP.EQ.0.)15,16
   15 DO 6 K=1,NN
      DISRAT(K)=DIST(1)/DIST(K)
      U(K+1)=U(K)+DISRAT(K)
    6 CONTINUE
      GO TO 18
   16 DO 17 K=1,NN
   17 U(K+1)=U(K)+DIST(K)
   18 LX=NN+1
      DO 7 J=1,NT
      KK=J-1
   12 CALL RANDU(IY,IY,V)
      Z=V*U(NNP1)
      DO 8 JJ=2,NNP1
      IF(Z.LE.U(JJ))9,8
    9 IF(J.EQ.1)GO TO 10
      DO 11 KL=1,KK
      IF(IIDVIS(KL).EQ.IDVIS(JJ-1))GO TO 8
   11 CONTINUE
   10 IIDVIS(J)=IDVIS(JJ-1)
      DDIST(J)=DIST(JJ-1)
      GO TO 7
    8 CONTINUE
    7 CONTINUE
C    HAVING FOUND THE TOWNS CHOSEN ACCORDING TO A PROBABILISTIC RULE CALL
C    THE SUBROUTINE CALC.
      CALL CALC(NUM,NT,HE,HN)
      RETURN
      END
```

```
      SUBROUTINE CALC(NUM,NT,HE,HN)
      COMMON ITD(1500),IE(1500),IN(1500),IPOP(1500),
     1IIDVIS(9),LIMPOP(7),LIMDIS(7,4),          DDIST(50)
      COMMON IY
      DIMENSION LOCPOS(27),LOCVIS(27)
      INTEGER HE,HN
C     COMPUTE AND WRITE OUT FOR EACH HOUSEHOLD THE NUMBER OF ALTERNATIVES
C     REJECTED IN EACH LOCATION TYPE AND THE LOCATION TYPES PATRONISED IN
C     THE ORDER OF PREFERENCE DEFINED BY THE PREFERENCE RULE.
      DO 2 LL=1,27
      LOCVIS(LL)=0
    2 LOCPOS(LL)=0
      DO 100 K=1,1342
      IF(IPOP(K).EQ.0)GO TO 100
      IDIST=IABS(HE-IE(K))+IABS(HN-IN(K))
      DO 20L=1,7
       IF(IPOP(K).LT.LIMPOP(L))21,20
   21 IF(L.EQ.1)22,24
   22 DO 23J=1,3
      IF(IDIST.LT.LIMDIS(L,J))222,23
   23 CONTINUE
      GO TO 100
   24 DO 27J=1,4
      IF(IDIST.LT.LIMDIS(L,J))222,27
   27 CONTINUE
      GO TO 100
  222 IF(L.EQ.1)13,14
   13 KK=J
      GO TO 12
   14 KK=((L-1)*4)+J-1
   12 DO 10 I=1,NT
      IF(IIDVIS(I).EQ.ITD(K))11,10
   11 LOCVIS(I)=KK
      GO TO 100
   10 CONTINUE
      LOCPOS(KK)=LOCPOS(KK)+1
      GO TO 100
   20 CONTINUE
  100 CONTINUE
      WRITE(1,1)NUM,HE,HN,NT,(LOCPOS(L),L=1,27),(LOCVIS(K),K=1,NT)
    1 FORMAT(1X,I2,3I3,2X,27I2,2X,(10I3))
      RETURN
      END
```

Program: TRANAD

For each triplet of location types, this program searches the entire sample and tests if each household makes independent implicit paired comparisons of one or two of the three possible pairs. A complete definition of independence is given in Chapter 4 (page 94). Using the independent paired comparisons of each pair in the triplet, the proportions are tested for transitivity, and a statistical test for the proportions' additivity is performed.

```
      PROGRAM TRANAD(INPUT,OUTPUT,TAPE1,TAPE5=INPUT,TAPE6=OUTPUT)
      DIMENSION LOCNM(27),LOCVIS(10),IWIN(60),ILOSE(60),PROX(60),
     1                NLL(3),SLL(3),F(3),ICHOSE(3),IND(3)
      REAL NLL
      JZ=25
  606 J25=JZ-2
      J26=JZ-1
      DO 611 J=1,J25
      JSEC=J+1
      DO 612 K=JSEC,J26
      JTHIR=K+1
      DO 613 L=JTHIR,JZ
      DO 700 KV=1,3
      NLL(KV)=0.
      SLL(KV)=0.
      F(KV)=0.
  700 CONTINUE
      REWIND 1
C     READ IN HOUSEHOLD≠S NUMBER OF LOCATION TYPES PATRONISED, LOCATION
C     TYPES AVAILABLE, AND LOCATION TYPES PATRONISED IN ORDER OF $
C     EXPENDITURE.
  616 READ(1,614)JB,(LOCNM(JJ),JJ=1,JZ),(LOCVIS(N),N=1,JB)
  614 FORMAT(9X,I3,2X,27I2,2X,(10I3))
      IF(EOF,1)633,617
  617 IF(JB.EQ.1)661,662
  661 ICHOSE(1)=LOCVIS(1)
C     TEST IF HOUSEHOLD≠S IMPLICIT PAIRED COMPARISONS FOR THIS TRIPLET
C     OF LOCATION TYPES(J,K AND L) ARE INDEPENDENT.
  673 IF(ICHOSE(1).EQ.J)618,619
  618 IF(LOCNM(K).GT.0.AND.LOCNM(L).EQ.0)624,625
  624 NLL(2)=NLL(2)+1.
      GO TO 616
  625 IF(LOCNM(K).EQ.0.AND.LOCNM(L).GT.0)626,616
  626 NLL(3)=NLL(3)+1.
      GO TO 616
  619 IF(ICHOSE(1).EQ.K)620,621
  620 IF(LOCNM(J).GT.0.AND.LOCNM(L).EQ.0)627,628
  627 NLL(2)=NLL(2)+1.
      SLL(2)=SLL(2)+1.
      GO TO 616
  628 IF(LOCNM(J).EQ.0.AND.LOCNM(L).GT.0)629,616
  629 NLL(1)=NLL(1)+1.
      GO TO 616
  621 IF(ICHOSE(1).EQ.L)622,616
  622 IF(LOCNM(J).GT.0.AND.LOCNM(K).EQ.0)630,631
  630 NLL(3)=NLL(3)+1.
      SLL(3)=SLL(3)+1.
      GO TO 616
  631 IF(LOCNM(J).EQ.0.AND.LOCNM(K).GT.0)632,616
  632 NLL(1)=NLL(1)+1.
      SLL(1)=SLL(1)+1.
```

```
      GO TO 616
  662 KQ=0
      DO 676 KS=1,3
      IND(KS)=0
  676 CONTINUE
  669 DO 663 MM=1,JB
      IF(MM.EQ.1)666,667
  667 NN=MM-1
      DO 668 LT=1,NN
      IF(LOCVIS(MM).EQ.LOCVIS(LT))663,668
  668 CONTINUE
  666 IF(LOCVIS(MM).EQ.J)664,665
  664 KQ=KQ+1
      ICHOSE(KQ)=J
      IND(1)=1
      GO TO 663
  665 IF(LOCVIS(MM).EQ.K)670,671
  670 KQ=KQ+1
      ICHOSE(KQ)=K
      IND(2)=1
      GO TO 663
  671 IF(LOCVIS(MM).EQ.L)672,663
  672 KQ=KQ+1
      ICHOSE(KQ)=L
      IND(3)=1
  663 CONTINUE
      IF(KQ.EQ.0)GO TO 616
      GO TO(673,674,675),KQ
  674 IF(IND(1).EQ.0)678,677
  678 NLL(1)=NLL(1)+1.
      IF(ICHOSE(1).EQ.L)679,680
  679 SLL(1)=SLL(1)+1.
  680 IF(LOCNM(J).EQ.0)GO TO 616
  681 IF(ICHOSE(2).EQ.L)682,683
  682 NLL(3)=NLL(3)+1.
      SLL(3)=SLL(3)+1.
      GO TO 616
  683 NLL(2)=NLL(2)+1.
      SLL(2)=SLL(2)+1.
      GO TO 616
  677 IF(IND(2).EQ.0)778,776
  778 NLL(3)=NLL(3)+1.
      IF(ICHOSE(1).EQ.L)779,780
  779 SLL(3)=SLL(3)+1.
  780 IF(LOCNM(K).EQ.0)GO TO 616
  781 IF(ICHOSE(2).EQ.L)782,783
  782 NLL(1)=NLL(1)+1.
      SLL(1)=SLL(1)+1.
      GO TO 616
  783 NLL(2)=NLL(2)+1.
      GO TO 616
  776 NLL(2)=NLL(2)+1.
      IF(ICHOSE(1).EQ.K)684,685
  684 SLL(2)=SLL(2)+1.
  685 IF(LOCNM(L).EQ.0)GO TO 616
      IF(ICHOSE(2).EQ.K)686,687
  686 NLL(1)=NLL(1)+1.
      GO TO 616
  687 NLL(3)=NLL(3)+1.
      GO TO 616
  675 IF(ICHOSE(1).EQ.L)688,689
  688 IF(ICHOSE(2).EQ.K)690,691
  690 NLL(1)=NLL(1)+1.
      SLL(1)=SLL(1)+1.
```

```
      NLL(2)=NLL(2)+1.
      SLL(2)=SLL(2)+1.
      GO TO 616
  691 NLL(3)=NLL(3)+1.
      SLL(3)=SLL(3)+1.
      NLL(2)=NLL(2)+1.
      GO TO 616
  689 IF(ICHOSE(1).EQ.K)692,693
  692 IF(ICHOSE(2).EQ.L)694,695
  694 NLL(1)=NLL(1)+1.
      NLL(3)=NLL(3)+1.
      SLL(3)=SLL(3)+1.
      GO TO 616
  695 NLL(2)=NLL(2)+1.
      SLL(2)=SLL(2)+1.
      NLL(3)=NLL(3)+1.
      GO TO 616
  693 IF(ICHOSE(2).EQ.L)696,701
  696 NLL(3)=NLL(3)+1.
      NLL(1)=NLL(1)+1.
      SLL(1)=SLL(1)+1.
      GO TO 616
  701 NLL(2)=NLL(2)+1.
      NLL(1)=NLL(1)+1.
      GO TO 616
  633 DO 640 KR=1,3
      IF(NLL(KR).EQ.0.)GO TO 613
      IF(SLL(KR).EQ.0.)634,635
  634 F(KR)=0.
      GO TO 640
  635 F(KR)=SLL(KR)/NLL(KR)
  640 CONTINUE
C     TEST IF TRIPLET OF LOCATION TYPES IS TRANSITIVELY RELATED.
  639 IF(F(1).GT..5.AND.F(2).GT..5.AND.F(3).LE..5.OR.F(1).LT..5.AND.F(2
     1.LT..5.AND.F(3).GE..5)GO TO 650
      IF(F(1).EQ..5)642,643
  642 IF(F(2).GT..5.AND.F(3).LT..5.OR.F(2).LT..5.AND.F(3).GT..5)650,652
  643 IF(F(2).EQ..5)644,645
  644 IF(F(1).GT..5.AND.F(3).LT..5.OR.F(1).LT..5.AND.F(3).GT..5)650,652
  645 IF(F(3).EQ..5)646,652
  646 IF(F(1).GT..5.AND.F(2).GT..5.OR.F(1).LT..5.AND.F(2).LT..5)650,652
  652 TRAN=4RTRAN
      GO TO 6500
  650 TRAN=6RINTRAN
 6500 PRINT 651,TRAN,(F(KQ),KQ=1,3),(NLL(KQ),KQ=1,3),L,K,J
  651 FORMAT(1X,R6,3F9.3,10X,3F5.0,10X,3I3)
C     TEST IF SAMPLE SIZE PERMITS ADDITIVITY TEST ON TRIPLET.
  647 T=NLL(3)-SLL(3)
      IF(SLL(3).GE.5..AND.T.GE.5.)653,613
  653 T=NLL(2)-SLL(2)
      IF(SLL(2).GE.5..AND.T.GE.5.)654,613
  654 T=NLL(1)-SLL(1)
      IF(SLL(1).GE.5..AND.T.GE.5.)655,613
  655 W=(SLL(1)/(NLL(1)-SLL(1)))*(SLL(2)/(NLL(2)-SLL(2)))
      X=W/(1.+W)
      FNUM=F(3)-X
      P=(F(3)+X)/2.
C     TEST FOR ADDITIVITY IN PROPORTIONS USING Z TEST.
      DENOM=SQRT(P*(1.-P)*(1./NLL(1)+1./NLL(2)+1./NLL(3)))
      Z=FNUM/DENOM
      IF(Z.GT.1.96.OR.Z.LT.-1.96)656,657
  656 ASTER=2R *
      GO TO 658
  657 ASTER=2RNS
  658 PRINT 659,ASTER,Z,X
  659 FORMAT(1H0,R2,F8.3,14X,F9.3)
  613 CONTINUE
  612 CONTINUE
  611 CONTINUE
      STOP
      END
```

BIBLIOGRAPHY

List of Abbreviations

AAAG Annals of the Association of American Geographers

JRS Journal of Regional Science

PPRSA Papers and Proceedings of the Regional Science Association

JAIP Journal of the American Institute of Planners

Abelson, R. P., and J. W. Tuckey. "Efficient conversion of nonmetric information into metric information", Proceedings American Statistical Association Meetings, Social Statistics Section, (1959), 226-230.

Abelson, R. P., and J. W. Tuckey. "Efficient utilization of non-numerical information in quantitative analysis: general theory and the case of simple order", Annals Mathematical Statistics, XXXIV, (1963), 1347-1369.

Alonso, William. Location and Land Use. Cambridge, Massachusetts: Harvard University Press, 1964.

Becker, G. M., M. H. De Groot and J. Marschak. "Probabilities of Choices among Very Similar Objects: an Experiment to Decide between Two Models", Behavioral Science, VIII, (1963), 306-311.

Berry, Brian J. L. Geography of Market Centers and Retail Distribution. Englewood Cliffs, New Jersey: Prentice-Hall, 1967.

Berry, Brian J. L., H. G. Barnum, and R. J. Tennant. "Retail Location and Consumer Behavior", PPRSA, IX, (1962), 65-106.

Brown, W. "The Judgement of Difference", University of California Publication in Psychology, I, No. 1, (1910).

Bucklin, Louis P. Shopping Patterns in an Urban Area. Berkeley: Institute of Business and Economic Research, University of California, 1967.

Bureau of Population and Economic Research, University of Virginia. "The Impact of a New Manufacturing Plant upon the Socioeconomic Characteristics and Travel Habits of the People in Charlotte County, Virginia", preliminary edition, University of Virginia, Charlottesville, 1951.

Carrol, J. Douglas and Howard B. Bevis. "Predicting Local Travel in Urban Regions", PPRSA, III, (1957), 183-197.

Christaller, Walter. Central Places in Southern Germany. Translated by Carlisle W. Baskin. Englewood Cliffs, New Jersey: Prentice-Hall, 1966.

Clarke, F. R. "Confidence Ratings, Second-Choice Responses, and Confusion Matrices in Intelligibility Tests", Journal of the Acoustical Society of America, XXXII, (1960), 35-46.

Coombs, Clyde H. A Theory of Data. New York: Wiley, 1964.

Curry, Leslie. "The Random Spatial Economy: An Exploration in Settlement Theory", AAAG, LIV, (1964), 138-146.

David, H. A. The Method of Paired Comparisons. London: Griffin, 1963.

Edwards, Ward. "The Theory of Decision Making", Psychological Bulletin, LI, (1954), 380-417.

Edwards, Ward. "Behavioral Decision Theory", Annual Review of Psychology, XII, (1961), 473-498.

Edwards, Ward and Amos Tversky, eds. Decision Making. Harmondsworth, England: Penguin Books, 1967.

Ewing, Gordon. "A Test to Assess the Significance of Differing Space Preferences in Consumer Spatial Choice Behavior", unpublished

M. A. thesis, McMaster University, Hamilton, Canada, 1968.

Fechner, G. Elements of Psychophysics. Translated by H. E. Adler; New York: Holt, Rinehart and Winston, 1966.

Gardner, J. "A Study of Neighborhood Travel Habits in Baltimore, Maryland", unpublished M. A. Thesis, Cornell University, 1949.

Garrison, William L. "Estimates of the Parameters of Spatial Interaction", PPRSA, II, (1956), 280-288.

Getis, Arthur. "A Theoretical and Empirical Enquiry into the Spatial Structure of Retail Activities", unpublished Ph.D. dissertation, University of Washington, Seattle, 1961.

Golledge, R. G. "Conceptualizing the Market Decision Process", JRS, VII, (1967), 239-258.

Goodchild, Michael F. "Local Structures in the Town Populations of Iowa", Geographical Analysis, I, (1969), 404-408.

Greeno, James G. Elementary Theoretical Psychology. Reading, Massachusetts: Addison-Wiley, 1968.

Guttman, Louis. "A general nonmetric technique for finding the smallest coordinate space for a configuration of points", Psychometrika, XXXIII, (1968), 469-506.

Hägerstrand, Torsten. "Aspects of the Spatial Structure of Social Communication and the Diffusion of Information", PPRSA, XVI, (1966), 27-42.

Hamburg, J. R. "Some Social and Economic Factors Related to Intra-City Movement", unpublished M. A. Thesis, Wayne State University, Detroit, 1957.

Harvey, David W. "Geographical Processes and the Analysis of Point Patterns", Transactions and Papers, Institute of British Geographers, XL, (1966), 81-95.

Huff, David L. "Towards a General Theory of Consumer Travel Behavior", unpublished doctoral dissertation, University of Washington, Seattle, 1959.

Huff, David L. "A Topographical Model of Consumer Space Preferences", PPRSA, VI, (1960), 159-173.

Huff, David L. "A Probabilistic Analysis of Consumer Spatial Behavior", in W. S. Decker, ed., Emerging Concepts in Marketing, Proceedings of the Winter Conference of the American Marketing Association, Chicago: 1962.

Iklé, Fred C. "Sociological Relationship of Traffic to Population and Distance", Traffic Quarterly, VIII, (1954), 123-136.

Isard, Walter. Location and Space-Economy. Cambridge, Massachusetts: M.I.T. Press, 1956.

Jonassen, C. T. The Shopping Center Versus Downtown. Columbus, Ohio: Bureau of Business Research, Ohio State University, 1955.

Kendall, Maurice G. Rank Correlation Methods. Second Edition; London: Griffin, 1955.

Kruskal, J. B. "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis", Psychometrika, XXIX, (1964a), 1-27.

Kruskal, J. B. "Nonmetric multidimensional scaling: a numerical method", Psychometrika, XXIX, (1964b), 115-129.

Lingoes, James C. "An IBM-7090 program for Guttman-Lingoes Smallest Space Analysis - I", Behavioral Science, X, (1965), 183-184.

Lingoes, James C. "New computer developments in pattern analysis and nonmetric techniques" in Uses of Computers in Psychological Research - The 1964 IBM Symposium of Statistics. Paris: Gauthier-Villars, 1966a, 1-22.

Lingoes, James C. "An IBM 7090 program for Guttman-Lingoes smallest space analysis - RI", Behavioral Science, XI, (1966b), 322.

Lingoes, James C. "A general nonparametric model for representing objects and attributes in a joint metric space", in J. C. Gardin, ed., Les compte-rendus de colloque international sur l'emploi des calculateurs en archeologie: problèmes sémiologiques et mathématiques. Marseille: Centre National de la Recherche Scientifique, 1969, (in press).

Lingoes, James C. Revised Guttman-Lingoes Programs - an annotated listing of the computer program G-L-SSA-I, 1970.

Lingoes, James C. and E. Roskam. "An empirical study of two multi-dimensional scaling algorithms", Multivariate Behavioral Research, V, (1970), (in press).

Lösch, August. Die räumliche Ordnung der Wirtschaft, translated by W. H. Woglom and W. F. Stolper as The Economics of Location. New Haven: Yale University Press, 1954.

Luce, R. Duncan. Individual Choice Behavior: a Theoretical Analysis. New York: Wiley, 1959.

Luce, R. Duncan and Patrick Suppes. "Preference, Utility and Subjective Probability", in R. Duncan Luce, R. R. Bush and E. Galanter, eds., Handbook of Mathematical Psychology, Volume III. New York: Wiley, 1965.

Marble, Duane F. "Transport Inputs at Urban Residential Sites", unpublished Ph.D. dissertation, University of Washington, Seattle, 1959.

Marble, Duane F. "Transport Inputs at Urban Residential Sties", PPRSA, V, (1959), 253-266.

Marble, Duane. "A Theoretical Exploration of Individual Travel Behavior", in W. L. Garrison and D. F. Marble, eds., Quantitative Geography, Part I: Economic and Cultural Topics. Evanston, Illinois: Department of Geography, Northwestern University, 1967.

Mayfield, Robert C. "The Range of a Central Good in the Indian Punjab", AAAG, LIII, (1963), 38-49.

McGee, V. E. "The multidimensional analysis of 'elastic' distances", British Journal of Mathematical and Statistical Psychology, XIX, (1966), 181-196.

Mertz, W. L. and Hamner, L. B. "A Study of Factors Related to Urban Travel", Public Roads, XXIX, (1957), 208-212.

Michelson, William. "An Empirical Analyses of Urban Environmental Preferences", JAIP, XXXII, (1966), 355-360.

Miller, G. A. "The Study of Intelligent Behavior", Annals of the Computation Laboratory of Harvard University, XXXI, (1962), 8.

Mitchell, Richard A. "An explanation of the expenditure pattern of a dispersed population", unpublished Ph.D. dissertation, State University of Iowa, Iowa City, 1964.

Moore, Eric G. "The Nature of Intra-Urban Migration and Some Relevant Research Strategies", Proceedings of the Association of American

Geographers, I, (1969), 113-116.

Murdie, Robert A. "Cultural Differences in Consumer Travel", Economic Geography, XLI, (1965), 211-233.

Peterson, George L. "A Model of Preference: Quantitative Analyses of the Perception of the Visual Appearance of Residential Neighborhoods", JRS, VII, (1967a), 19-31.

Peterson, George L. "Measuring Visual Preferences of Residential Neighborhoods", Ekistics, XXIII, (1967b), 169-173.

Ray, D. Michael. "Cultural Differences in Consumer Travel Behavior in Eastern Ontario", Canadian Geographer, XI, (1967), 143-156.

Roskam, E. E. C. I. "A comparison of principles for algorithm construction in nonmetric scaling", Michigan Mathematical Psychology Program, Technical Report No. MMPP69-2, 1969.

Roskam, E. and James C. Lingoes. "MINISSA-I: A Fortran IV(G) program for the smallest space analysis of square symmetric matrices. Behavioral Science, XV, (1970), (in press).

Rushton, Gerard. "Spatial Competition for the Supply of Goods and Services to the Iowa Dispersed Population", Iowa Business Digest, XXXV, (1964), 3-8.

Rushton, Gerard. Spatial Pattern of Grocery Purchases by the Iowa Rural Population. Iowa City: Bureau of Business and Economic Research, University of Iowa, 1966.

Rushton, Gerard. "The Scaling of Locational Preferences", in Kevin Cox and R. G. Golledge, eds., Problems of Spatial Behavior: A Symposium. Evanston, Illinois: Department of Geography, Northwestern University, 1969a.

Rushton, Gerard. "Temporal Changes in Space Preference Structures", PAAG, I, (1969b), 129-132.

Rushton, Gerard. "Postulates of Central Place Theory and the Properties of Central Place Systems", Geographical Analysis, II, (1970), forthcoming.

Russell, Bertrand. Analysis of Mind, London: G. Allen and Unwin, 1921.

Shepard, Roger N. "The analysis of proximities: multidimensional scaling with an unknown distance function: I", Psychometrika, XXVII, (1962a), 125-140.

Shepard, Roger N. "The analysis of proximities: multidimensional scaling with an unknown distance function: II", Psychometrika, XXVII, (1962b), 219-246.

Shepard, Roger N. "On Subjectively Optimum Selections among Multi-attribute Alternatives" in M. W. Shelley and G. L. Bryan, eds., Human Judgements and Optimality. New York: Wiley, 1964.

Shepard, Roger N. "Metric structures in ordinal data", Journal of Mathematical Psychology, III, (1966), 287-315.

Sherman, Charles R. and Forrest W. Young. "Nonmetric multidimensional scaling: a Monte Carlo study", Proceedings, 76th Annual Convention, American Psychological Association, 1968, 207-208.

Siegel, Sidney. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Stafford, H. "The Functional Bases for Small Iowa Towns", Economic Geography, XXXIX, (1963), 165-175.

Thomas, Edwin N. "Some Comments on the Functional Bases for Small Iowa Towns", Iowa Business Digest, XXXI, (1960), 10-16.

Thomas, Edwin N., R. A. Mitchell and D. A. Blome. "The Spatial Behavior of a Dispersed Non-farm Population", PPRSA, IX, (1962), 107-133.

von Thünen, Johann H. Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalekonomie. Translated by Peter Hall as von Thünen's Isolated State. London: Pergamon Press, 1966.

Tocher, K. D. "Extension of the Neyman-Pearson theory of tests to discontinuous variates", Biometrika, XXXVII, (1950), 130-144.

Torgerson, Warren S. Theory and Methods of Scaling. New York: Wiley, 1958.

Tucker, Ledyard R. "Intra-individual and Inter-individual Multidimensionality", in H. Gulliksen and S. Messick, eds., Psychological Scaling: Theory and Applications. New York: Wiley, 1960.

Tversky, Amos. "Additivity, utility and subjective probability", Journal of Mathematical Psychology, IV, (1967), 175-202.

Weber, Alfred. Über den Standort der Industrien, translated by C. J. Friedrich as Alfred Weber's Theory of the Location of Industries. Chicago: University of Chicago Press, 1928.

Wolpert, Julian. "Behavioral Aspects of the Decision to Migrate", PPRSA, XV, (1965), 159-169.

Yntema, D. B. and W. S. Torgerson. "Man-Computer Cooperation in Decisions Requiring Common Sense", Institute of Radio Engineers, Transactions on Human Factors in Electronics, HFE-2, (1961), 20-26.

Young, F. W. and W. S. Torgerson. "TORSCA: A Fortran IV program for Shepard-Kruskal multidimensional scaling analysis", Behavioral Science, XII, (1967), 498.