CONSUMER SPACE PREFERENCES

A TEST TO ASSESS THE SIGNIFICANCE OF DIFFERING SPACE PREFERENCES IN CONSUMER SPATIAL CHOICE BEHAVIOR

By

GORDON ORR EWING, M. A.

A Thesis
Submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
For the Degree
Master of Arts

McMaster University

May, 1968

MASTER OF ARTS (1968) McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: A Test to Assess the Significance of Differing Space Preferences in Consumer Spatial Choice Behavior

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

SUPERVISOR: Professor G. Rushton

NUMBER OF PAGES:

SCOPE AND CONTENTS: The study seeks to explain a dispersed population's spatial choices of urban places for retail expenditures. Specifically, it tests the previously untested hypothesis that consumers evaluate the same spatial opportunities differently, with different evaluations of spatial opportunities being defined as differing space preferences. Using a model, which has the constraint that there are no differences in consumer space preferences, its predictive power is increased significantly when that constraint is relaxed. On this basis, the hypothesis that there are differing space preferences is accepted. In addition, the hypothesis is tested that differences in space preferences are related to

differences in the socioeconomic characteristics of consumers. No relationship is found, and an alternative analysis is suggested in the light of weakness in the existing test. Finally, the hypothesis is accepted that the predictive power of the model is least where a household has to choose between towns ranked closely by the model.

Acknowledgements

The writer wishes to thank Dr. Gerard Rushton, who, as supervisor, devoted much time to providing invaluable comment and criticism at all stages of the preparation of this thesis.

Table of Contents

Table of Contents (Continued)

List of Illustrations

List of Illustrations (Continued)

List of Illustrations (Continued)

List of Tables

List of Tables (Continued)

INTRODUCTION

The Purpose

Consumers who are faced with the same set of spatial alternatives, nevertheless make different spatial choices. The purpose of this study is to test the hypothesis that these differences are functionally related to personal characteristics of the consumer. Called in question, therefore is the basic premise of several previous studies[1], which have postulated that consumer spatial choice behavior could be explained without reference to the consumer's personal traits. Deficiencies in the explanation of consumer spatial choice behavior found in these studies, together with statements in the literature regarding interpersonal variations in spatial choice behavior, lend weight to the purpose of this enquiry. It is hypothesized that models of spatial choice behavior will have greater predictive power, if they include measures of the personal characteristics of the consumer.

[1] Huff, D. L., "A Probabilistic Analysis of Consumer Spatial Behavior", Emerging Concepts in Marketing; Proceedings of the Winter Conference of the American Marketing Association, W. S. Decker (ed.), December, 1962.
Mitchell, R. A., "An Explanation of the Expenditure Pattern of a Dispersed Population", unpublished Ph. D. dissertation, State University of Iowa, 1964.
Berry, B. J. L., Barnum, H. G., and Tennant, R. J., "Retail Location and Consumer Behavior", Papers and Proceedings of the Regional Science Association, Volume 9, 1962, pp. 65-106.

The Study and Behaviorism

It is to be noted that "behavior" is central to the study's hypothesis. In particular, the emphasis is placed upon how people actually do or can behave. This point is made in a study by Siegel, Siegel and Andrews who stress that,

> "analyses are directed to the ways in which people actually behave, not to how they say they behave or would behave nor to how they might expect others to behave. In our judgement, the hypothesis of maximization of expected utility[2] can be given a fair test only by research in the behaviorist tradition"[3].

The argument for behavioristic rather than motivational research has been the concern of psychologists in particular, who, in seeking to develop their discipline as an objective science, have had to question the validity of motivational research. Miller suggests that,

> "....behaviorism, in America, gradually replaced mentalism. 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

[2] Consumer spatial choice behavior is assumed to reflect the attempt to maximize the utility of a given choice amongst alternative spatial opportunities for retail expenditure.

[3] Siegel, S., Siegel, A. E., and Andrews, J., "Choice, Strategy and Utility", McGraw-Hill, New York, 1964, p. 19.

> 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"[4].

Thus, it is unreasonable to assume that consumers are aware of the motivation of an act. Indeed, the consumer's awareness of a conscious motivation does not, of necessity, reduce the likelihood of the existence of a regular behavior pattern reflecting responses to consumer needs. For these reasons, it is argued that behavior is a more valid basis for research into consumer spatial behavior, than is an enquiry into motivation behind behavior.

The Study and Central Place Theory[5]

Central place studies are essentially concerned to explain the relationship between urban places and the dispersed population which they serve. This study focusses on that population's spatial choices of central places as places providing required goods and services. Central Place theory, however, is based upon certain assumptions about spatial choice behavior, one of which is that the consumer goes to the nearest place offering a desired good. Implicit in this assumption is the inference that all

[4] Miller, G. A., "The Study of Intelligent Behavior", Annals of the Computation Laboratory of Harvard University, Volume 31, 1962, p. 8.

[5] Most fully propounded by Walter Christaller in "Central Places in Southern Germany", translated by Carlisle W. Baskin, Prentice-Hall, 1966.

consumers have the same regularity in their spatial choice behavior. That inference is questioned in this study, which tests for the existence of a pattern in consumer spatial behavior, which is functionally related to consumer, as well as spatial, variables.

Since central places exist in response to the needs of consumers, whose spatial choices affect their distribution and functional diversity, an understanding of consumer spatial choice behavior is particularly important in explaining the spatial distribution and functional diversity of central places. Hence, an understanding of the underlying factors influencing spatial choice behavior would contribute to the explanation of the size and spacing of centers.

CHAPTER 1

The Problem of Consumer Rankings of Places of Purchase

1.1 The Problem

The basic premise tested in this study is that there are significant inter-personal differences in consumers' rankings of alternative spatial opportunities for the allocation of retail expenditures. These rankings, regarded as representing consumers' preferences, are assumed to be revealed in the observed spatial choice behavior of consumers. It would be possible to distinguish from spatial behavior at least four possible alternative types of preference rankings. These are:

a) that each consumer has a unique preference ranking, which may or may not be predictable in terms of certain intrinsic characteristics of the consumer,

b) that there is a uniform regularity in consumer preference rankings, such that if this regularity could be discerned and adequately described, it would always be possible to predict consumer spatial behavior,

c) that there are various sets of preference rankings, which either differ from one another randomly, or reveal a degree of regularity between one another, with some random inconsistencies between them: or there are

various preference rankings unrelated to any observable physical, social or psychological characteristic of the consumer, and of his environment, and

d) that there are various sets of preference rankings, which have an observable regularity functionally related to some measurable characteristic of the individual and of his environment.

The testing of the fourth alternative forms the nexus of this work for several reasons. The first alternative can be discounted in view of the partial success of existing models of consumer spatial behavior, which implies some degree of consistency amongst consumer preference rankings. The second alternative suggests the existence of a causal law which can be rigorously applied to every example of consumer spatial choice. Such a high degree of regularity in human behavior is considered unlikely. The third alternative is threefold. Firstly, wholly random preference rankings of spatial opportunities are not indicated in existing studies. Secondly, no immediate means are apparent to elucidate preference rankings which are unrelated to any observable consumer traits. Thirdly, it could be reasonably hypothesized that there exists a regularity in consumer preference rankings, which is not, however, a complete regularity, but rather has random inconsistencies within it. This idea would appear to have

most scope for development, once the fourth alternative has been tested.

The above reasons, together with the weight of previous speculations and related literature, constitute the justification for selecting the fourth alternative as the most deserving of study. That is, there are hypothesized to be various sets of preference rankings, which are functionally related to some measurable characteristics of the consumer and his environment.

1.2 Space Preferences as Discussed in the Literature

In the literature, the term "space preferences" has been used to describe the propensity of individuals for differing amounts of, or forms of, spatial interaction. Abstractly expressed, if spatial interaction is defined to have either positive or negative utility, then, in any individual's spatial behavior, it is hypothesized that there is demonstrated a certain need, or desire, for spatial interaction, be it positive or negative. Isard best exemplifies the notion when he says,

> "Psychologists and sociologists, whether speaking of a gregarious instinct or of acquired behavior patterns or of both, have emphasized the social nature of man and his propensity to associate with groups of various sorts. One can reason that such a propensity.....is a manifestation of a positive space preference.....It should be stressed that not all individuals

> need have a positive space preference. There are hermits. They exhibit negative space preference....."[1]

In this study, space preferences are defined as the propensity of individuals for differing forms of, rather than differing amounts of, spatial interaction, even though similarly located with respect to alternative spatial opportunities. The emphasis, unlike Isard's example, is upon interactions between consumers of retail services and central places, rather than upon inter-personal interactions. Used in the singular, space preference refers to the propensity for a given form of spatial interaction. In terms more specific to this study, this propensity can be restated as a consumer's mental ranking of all conceivable spatial locations where interaction is possible.[2] Such a ranking is revealed, partially at least, in the consumer's spatial choice behavior. Inevitably, any ranking observed can only be a ranking of actual alternative spatial locations, and not of all conceivable spatial locations.

The necessity for this kind of study is emphasized, again by Isard, when he says,

> "We especially need to probe deeply

[1] Isard, W., "Location and Space-Economy", Regional Science Studies Series, Number 1, Cambridge, Mass., 1962, p. 84.

[2] In the present context, spatial locations relevant to retail consumer expenditure are urban places with retailing facilities.

> into space preferences, i.e., into man's propensity for intricate forms and patterns of herd existence and into the socio-psychological and biological forces which together with economic and other forces govern the spatial patterns of population settlement"[3].

Such a need, it is argued, is equally strong with respect to the dynamic patterns of human movements in space, be they day-to-day, or over any longer time span.

More specific references to space preferences, in which patterns of spatial behavior are thought of as being related to socioeconomic variables, are to be found in the works of Michelson[4], Getis[5], Marble[6], Huff[7], and Malm, Olsson and Wärneryd[8]. Getis suggests that,

> "In general, the studies show that locational variables do not seem to

[3] Isard, W., op. cit., p. 287.

[4] Michelson, W., "An Empirical Analysis of Urban Environmental Preferences", Journal of the American Institute of Planners, Volume 32, Number 6, November, 1966, pp. 355 - 360.

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

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

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

[8] Malm, R., Olsson, G., Wärneryd, O., "Approaches to Simulations of Urban Growth", Geografiska Annaler, Volume 48B, Number 1, 1966.

> be any more important than socioeconomic characteristics in influencing consumer behavior"[9].

Marble says of general location theory 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"[10].

Malm, Olsson and Wärneryd go so far as to say that,

> "It is well verified that the distance decay function" [in a model of spatial interaction] "varies both with the hierarchical order of the interacting places, and with the demographical and socioeconomic characteristics of the interactors. Further, contacts tend to follow traditional, well-established channels.... Generally, such deviations from symmetrical patterns are termed 'space preferences'"[11].

Such a variety of sources pointing to the existence of space preferences in several fields of spatial behavior, would seem to argue strongly for the present study's hypothesis. However, from examples of empirical works to be discussed later, it will be shown that this work is not a repetition of any previous one, but seeks to examine spatial behavior in a significantly different and, hopefully, more meaningful fashion.

[9] Getis, A., op. cit., pp. 135 - 136.

[10] Marble, D. F., op. cit., p. 22.

[11] Malm, B., Olsson, G., Wärneryd, O., op. cit., p. 12.

1.3 The Basis of the Space Preference Hypothesis and Others

The actual analysis entails testing this hypothesis against the predictive results of a normative spatial allocation model developed by Rushton[12]. "The purpose of this model is to predict the particular town which any given farm household in Iowa will select as its maximum grocery purchase town"[13]. On the premise that the critical variables influencing this selection are the population of, and distance to alternative central places from each household, the model correctly predicts actual behavior for 65% of the sample households examined (see Appendix 1). Given this statistic, it is hypothesized that a significant part of the incorrect prediction is a consequence of not including any variables which might take account of space preferences. Space preferences, by definition, would account for similarly located households nevertheless behaving differently in space.

There are, however, alternative possible explanations of the model's partially incorrect prediction (figure 1.1), only some of which will be pursued, for reasons to be given. The first to be

12 Rushton, G., "Spatial Pattern of Grocery Purchases by the Iowa Rural Population", Studies in Business and Economics, New Series Number 9, University of Iowa, Iowa City, 1966.

13 Rushton, G., op. cit., pp. 27 - 28.

FIGURE 1.1

Possible Sources of Error in an Indifference Surface Model of Consumer Spatial Behavior.

Model

Explained Population		Unexplained Population: Residual		
1) Other variables necessary in model.	2) Inability of model to handle complex sets of alternative opportunities.	3) Existing variables need redefinition.	4) Incorrect functional relationship assumed.	5) Statistical techniques inappropriate

tested, as already described, is that inaccurate prediction is a function of the absence of certain "personal" variables in the model, i.e. that behavior varies in an orderly fashion as a function both of varying spatial situations and of varying intrinsic consumer characteristics.

It is also hypothesized that, in addition to or as an alternative to the above possible explanation, inaccurate prediction is a function of the model's inability to "realistically"[14] discriminate between spatial alternatives in complex choice situations. This itself may be the result of the inadequacy of the present technique for deriving the indifference surface from revealed spatial behavior.

[14] By "realistic" is meant "in the manner in which people actually do something".

The reason for restricting the problem to the testing of the above hypotheses is that these seemed on intuitive grounds to hold most of success. With respect to the other possible sources of error in Figure 1.1, it is arguable that their contribution to error is minimal. The existing variables of town size and distance have been widely employed in geographic studies as surrogates of town attractiveness and the friction of space. Distance is commonly regarded as being monotonically related to the friction of space. For example, in the literature pertaining to spatial aspects of marketing and retailing and studies using gravity models, a distance measure is invariably used as a surrogate for spatial friction. Similarly, studies have shown that the attractiveness of a central place for retail purchases is closely linked to the variety and amount of goods and services offered at the place[15], and that this in turn is highly correlated with town size[16].

[15] Christaller, W., op. cit.
Baumol, W. J., and Ide, E. A., "Variety in Retailing", Management Science, Volume 3, 1956, pp. 93-101.

[16] Christaller, W., op. cit.
Stafford, H., "The Functional Bases for Small Iowa Towns", Economic Geography, Volume 39, No. 2, 1963, pp. 165-175.
Thomas, E. N., "Some Comments on the Functional Bases for Small Iowa Towns", Iowa Business Digest, Volume 31, No. 2, Bureau of Business and Economic Research, University of Iowa, Iowa City, 1960

With reference to the suggestion that incorrect functional relationships may have been assumed, it seems from the model's correct prediction of 65% of spatial choices, that a reasonably accurate functional relationship has been demonstrated between spatial choice behavior and town attractiveness as measured by town size, together with the friction of space measured by distance.

The validity of the technique used to calculate the indifference surface (described in appendix 1) is discussed later.

The rationale behind these hypotheses is the quest to make more lawful statements about consumer spatial choice behavior, so enabling an improvement upon the predictive power of Rushton's indifference surface model. An understanding of the form of consumer space preferences would also be a useful key to the better comprehension of the form of the spatial economic system, developed in response to consumer needs and behavioral patterns.

CHAPTER 2

Models of Consumer Spatial Behavior- A Review

The development and testing of models of consumer spatial behavior is a relatively recent phenomenon in the history of geographical and marketing literature. Perhaps as a result of existing models being pioneers, the notion of space preference has never been incorporated in any of them, nor have adequate tests been made to assess its hypothesized significance. Certainly, it is not an uncommon hope amongst model-makers, that behavior can be explained with reference to a minimum of variables. However, the lack of power of existing models in explaining spatial behavior points up a clear need to examine other variables as possible additional explainers of behavior. Inevitably, too, such weaknesses call in question the basic methodology of the models. Therefore, it would be fruitful to consider the qualities and limitations of previous models of consumer spatial behavior, in the hope that directions of possible improvement might be highlighted, and that the present work might be viewed within the framework of earlier efforts.

2.1 The Regression Model

Essentially, regression analysis seeks to describe

a mathematical relationship between one dependent variable and a given number of independent variables, as expressed in the equation,

$$y = a + bx_1 + cx_2 + \ldots + px_n, \qquad (2.1)$$

where y = the dependent variable,

a = a constant term,

$x_1, x_2, \ldots x_n$ = n independent variables, and

$b, c, \ldots p$ = empirically derived multipliers of the relevant independent variables.

For each observation of the dependent variable there is an observation for each of the independent variables in the data. The functional relationship between the dependent and independent variables is shown when values of a, b, c, ... p are obtained, such that the sum of terms on the right-hand side of equation 2.1 equals the observed value of the dependent variable, for every set of observations.

The relevance of such a technique to the prediction of consumer spatial choice behavior was investigated by Mitchell[1]. Using as data the dollar retail expenditures of a sample of farm and non-farm households in Iowa, he attempted to predict, through a multiple regression model, the expenditure of each household (the dependent variable) in every town it patronised. In his initial argument he,

[1] Mitchell, R. A., "An Explanation of the Expenditure Pattern of a Dispersed Population", unpublished Ph. D. disseration, State University of Iowa, 1964.

> ".... hypothesized that the spatial allocations or transactions of a dispersed population can be explained by three broad categories of variables, i.e., $T = f(x_1, x_2, \ldots x_m, y_1, y_2, \ldots y_n, z_1, z_2, \ldots z_p)$, where T is the amount of expenditures executed in a given town by an individual farm household; the x variables represent spatial or distance factors; the y variables relate to the characteristics of the farm and household; and the z variables represent characteristics of the town"[2].

In the initial multiple regression model, there are four independent variables, one each for measures of distance and town centrality, and two to describe household characteristics, namely household size and income. The latter two, however, are used as consumption functions, since the dependent variable is expressed in dollars. In fact it can be argued that Mitchell is adding an unnecessary dimension to the problem of explaining spatial allocation of expenditures. For the critical purpose of such a study is to explain consumer rankings of alternative central places and, if possible, to ascribe some cardinal measure to this ranking. With such knowledge, it is then possible to predict both the towns with which a household interacts and the relative proportion of expenditure in each. Information on absolute dollar expenditures adds nothing to the model's explanatory power, and indeed carries over to the economist's problem of consumption functions.

[2] Mitchell, R. A., op. cit., p. 9.

Furthermore, a test of space preferences by using the two household variables proves to be impossible. Were it not for the "absolute dollar expenditure" nature of the dependent variable, the two household variables could be dropped from the initial model and the proportion of explained behavior, before and after their inclusion, compared. However, the need to know household income and size in order to predict absolute dollar expenditures eliminates the possibility of a space preference investigation. Thus, the model must be criticized for combining two independent problems and so denying the possibility of examining the relationship of household variables to space preferences.

The initial model explains only 35% of observed variation. The subsequent addition of an "intervening opportunity" variable and three dichotomous variables ("place of work", "county seat town", and "nearest place of purchase") increases explanation by only 4%.

As an explanation of the low predictive power of the model, he employs the argument of "irrational" behavior in consumer spatial behavior. Such behavior is defined as "an expenditure made in a town when there is a town of equal or greater size that is closer." Out of 633 expenditures, 211 are "irrational" by this definition. The exclusion of these "irrational" households from the data, allows the

proportion of explained variation to be increased to 47%. However, it seems unlikely that such an a priori definition of rationality can be empirically verified. Indeed there may be grounds for arguing that such spatial behavior may be entirely consistant with rational behavior. However, as a statistical model has no purpose to explain the reasoning behind behavior, but simply describes the "how" of behavior, the argument over "rational" behavior is largely irrelevant.

A more reasonable explanation proferred to account for the model's limitations, though not developed, is that,

> "...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"[3].

In light of such a remark, it seems unprofitable to pursue the idea that different explanatory regression equations could be developed for the spatial behavior of consumers with different space preferences. Furthermore, regression models only consider actual interactions, thus neglecting the significance of foregone opportunities for interaction. If an adequate descriptive model of consumer space preferences is sought, the model must be able to describe how consumers make trade-offs between alternatives. Such is

[3] Mitchell, R. A., op. cit., p. 85.

impossible in a model where only actual interactions are observed. An example of this weakness is provided by comparing the spatial behavior of two groups, one of which travels further for a given commodity than the other. Regression models describing these patterns would tell us nothing about the "willingness" of households in these groups to travel, nor about the way people rank alternative centers available to them. Thus, if one group is favorably located with respect to a large retail outlet or city, the regression model's (albeit accurate) description of differing spatial behavior patterns for the two groups, might convey a conclusion about their respective space preferences at variance with reality.

2.2 The Gravity Model

The concept of a gravity model of human interaction was first related to retailing by Reilly in 1931[4]. He postulated that for any two cities competing for retail trade, the point of equilibrium on the line between them, where drawing power is equal, will be described by the equation,

$$\frac{P_i}{d_{xi}^2} = \frac{P_j}{d_{xj}^2}, \qquad (2.2)$$

[4] Reilly, W. T., "The Law of Retail Gravitation", The Knickerbocker Press, New York, 1931.

where P_i, P_j = the population of cities i and j, respectively, and

d_{xi} = the distance from city i to point x, where x has that empirically determined value such that the expression is true.

The gravity model with the general form,

$$I_{ij} = \frac{f(P_i, P_j)}{f(D_{ij})}, \qquad (2.3)$$

where I_{ij} = the interaction between center i and center j,

P_i, P_j = the population of centers i and j respectively, and

D_{ij} = the distance between centers i and j,

has since been applied in several fields of human geography, but its direct application in a more complex model of consumer spatial behavior than Reilly's, did not appear until 1962[5]. In this paper Huff tests the postulate, implicit in the gravity model, that consumer spatial behavior can be lawfully described in terms of two environmental parameters, namely, the size of the shopping center in square feet of selling area and travel time to it. The gravity equation developed is as follows:-

$$P_{ij} = \frac{u_{ij}}{\sum_{j=1}^{n} u_{ij}} = \frac{\frac{S_j}{T_{ij}^{\lambda}}}{\sum_{j=1}^{n} \frac{S_j}{T_{ij}^{\lambda}}} \qquad (2.4)$$

[5] Huff, D. L., "A Probabilistic Analysis of Consumer Spatial Behavior", Emerging Concepts in Marketing, Proceedings of the Winter Conference of the American Marketing Association, W. S. Decker (ed.), December, 1962.

where P_{ij} = the probability of a consumer at a given point of origin i traveling to a given shopping center j;

S_j = the size of a shopping center j;

T_{ij} = the travel time involved in getting from a consumer's travel base i to shopping center j;

λ = a parameter which is to be estimated empirically to reflect the effect of travel time on various kinds of shopping trips"[6]; and,

u_{ij} = the utility of an interaction by consumer at travel base i with j^{th} shopping center.

This says, in effect, that the probability (P) of a consumer at point i interacting with shopping center j, equals the ratio of j's utility to the sum of the utilities of all shopping centers being considered. Huff defines utility as being a direct function of the size of a shopping center, and inversely related to some power of its distance from the consumer.

Furthermore, implicit in the model is a definition of <u>cardinal</u> utility, as one center's utility to the consumer, expressed as a ratio of the sum of the utilities of all centers (see equation 2.4). From this ratio is derived a measure of the shopping center's cardinal utility to the consumer in the form of a probability statement of interaction with that center, i.e.,

[6] Huff, D. L., op. cit., p. 445.

$$P_{ij} = \frac{u_{ij}}{\sum_{j=1}^{n} u_{ij}} \quad ; \quad \text{hence} \quad \sum_{j=1}^{n} P_{ij} = 1 \qquad (2.5)$$

Thus the model not only attempts to reveal consumer ordinal preference rankings of spatial opportunities, but also tries to predict the degree to which one center is preferred to another, i.e. it predicts preference rankings on a cardinal scale.

The model is not designed, however, to take into account the possible existence of space preferences, i.e. it postulates a uniform preference ranking of spatial situations for all consumers. Huff specifically indicates[7] that a consumer, given a sufficient number of choice situations, will not choose one center exclusively, but will choose centers with a frequency related to the calculated probability of an interaction between neighborhood i and center j (P_{ij}). Each consumer, therefore, is predicted to behave according to the same "rule", i.e. the frequency of interactions of all consumers with a given center is made equal.

However, despite Huff's constraint about similar patterns of behavior for all consumers, it can be logically argued that the probabilities can be regarded as statements of differing space preferences. Each probability of inter-

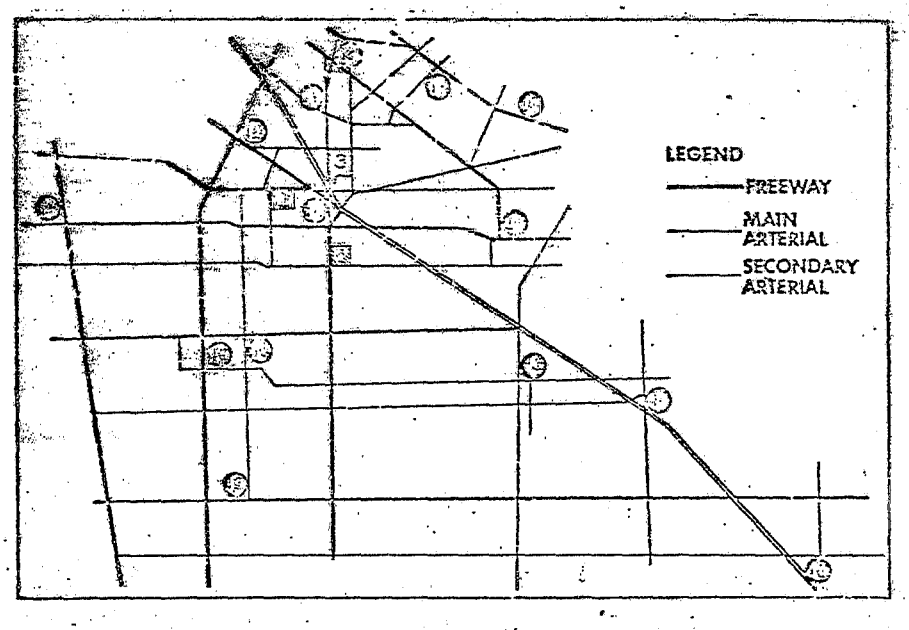
[7] Huff, D. L., op. cit., p. 446.

action can be regarded as a statement of the proportion of consumers who will always choose that particular center. In that case, there would be as many sets of space preferences predicted as there were probabilities. It is questionable whether this line of reasoning is defensible, but nevertheless the ambiguity of the model's relation to space preferences exists.

In addition, the gravity model, per se, as well as Huff's method of testing it, has certain demonstrable limitations, when applied to consumer spatial behavior. The spatial arrangement of the array of shopping centers and the three neighborhoods for which Huff attempts to explain spatial behavior are depicted in figure 2.1. For each neighborhood, Huff calculates the probability (P_{ij}, for j=1, n) of a consumer located in it, interacting with each shopping center. Given the number of consumers living in neighborhood i (C_i), it is a simple matter to calculate the number of consumers predicted to patronize each center ($P_{ij}.C_i$, where j=1, 2,...n). P_{ij} is derived using equation 2.4, and is calculated for every neighborhood-shopping center combination. In practice, the value of λ is varied systematically until that value of P_{ij} for each shopping center is reached, such that the expected frequency of interaction between neighborhood and center most closely approximates the observed frequency of interaction (see table 2.1 for λ values and table 2.2 for comparison of

FIGURE 2.1

Geographical Relationship of the
Shopping Centers to Sample Neighborhoods



Source: Huff, D. L., op. cit., p. 451

TABLE 2.1

Parameter Estimates for Each Neighborhood with Respect to Clothing and Furniture Purchases

Commodity	Neighborhood	Estimate of λ	Correlation Coefficient
Clothing	1	2.812	.99
Clothing	2	2.604	.98
Clothing	3	3.779	.96
Furniture	1	2.523	.99
Furniture	2	3.115	.94
Furniture	3	3.331	.96

Source: Huff, D.L., op. cit., p. 456

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 Centers

Shopping Center	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	2.80	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., op. cit., p. 454

observed and expected frequencies).

In fact the technique used to obtain this closest possible fit of observed and expected frequencies is remarkably analogous to the "sum of least squares fit" method in regression analysis. This analogy is also a clue to the limitations of a gravity model of consumer spatial behavior. Insofar as $\sum_{j=1}^{n} \frac{S_j}{T_{ij}^{\lambda}}$ has the same value in the calculation of P_{ij} for every shopping center, the equation 2.4 can be expressed as,

$$P_{ij} = \frac{\frac{S_j}{T_{ij}^{\lambda}}}{m}, \qquad (2.6)$$

where m is a constant, equal to the sum of the utilities of all shopping centers with respect to neighborhood i. Therefore,

$$mP_{ij} = \frac{S_j}{T_{ij}^{\lambda}} \quad . \qquad (2.7)$$

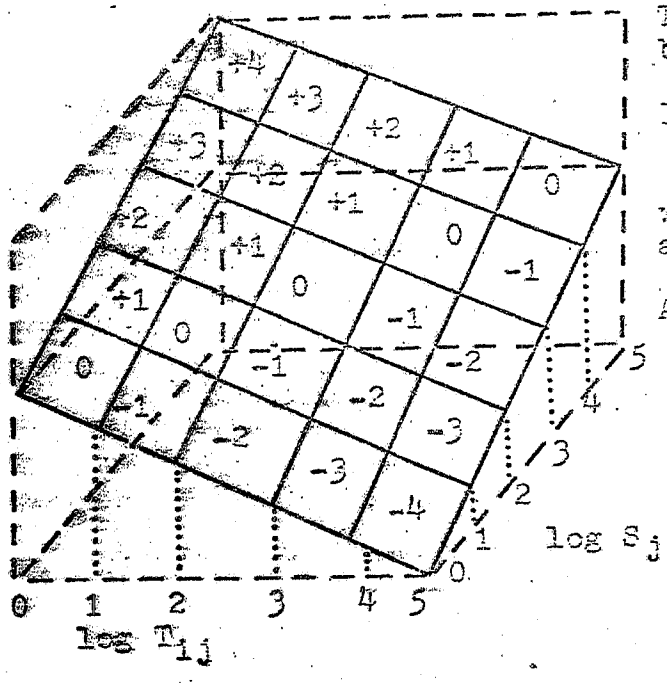
Logarithmic transformation of equation 2.7 produces the following:-

$$\log m + \log P_{ij} = \log S_j - \lambda \log T_{ij} , \qquad (2.8)$$

or, $$\log P_{ij} = - \log m + \log S_j - \lambda \log T_{ij} . \qquad (2.9)$$

Equation 2.9 has the form of a multiple regression equation with two independent variables. Using this transformation of the gravity model, and making λ equal to unity, the model would take the form shown in figure 2.2, and probably would have limited power of prediction. If, as in Huff's work, λ

FIGURE 2.2



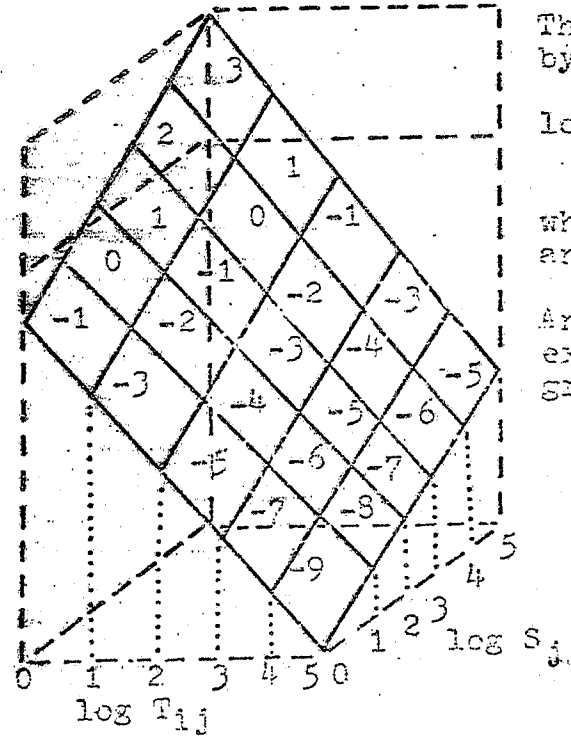
The plane described by the equation,

$$\log P_{ij} = -\log m + \log S_j - \lambda \log T_{ij},$$

where $-\log m = 0$,
and $\lambda = 1$.

Angle of plane = 45°.

FIGURE 2.3



The plane described by the equation,

$$\log P_{ij} = -\log m + \log S_j - \lambda \log T_{ij},$$

where $-\log m = 0$,
and $\lambda = 2$.

Angle of plane always exceeds 45°, if λ greater than 1.

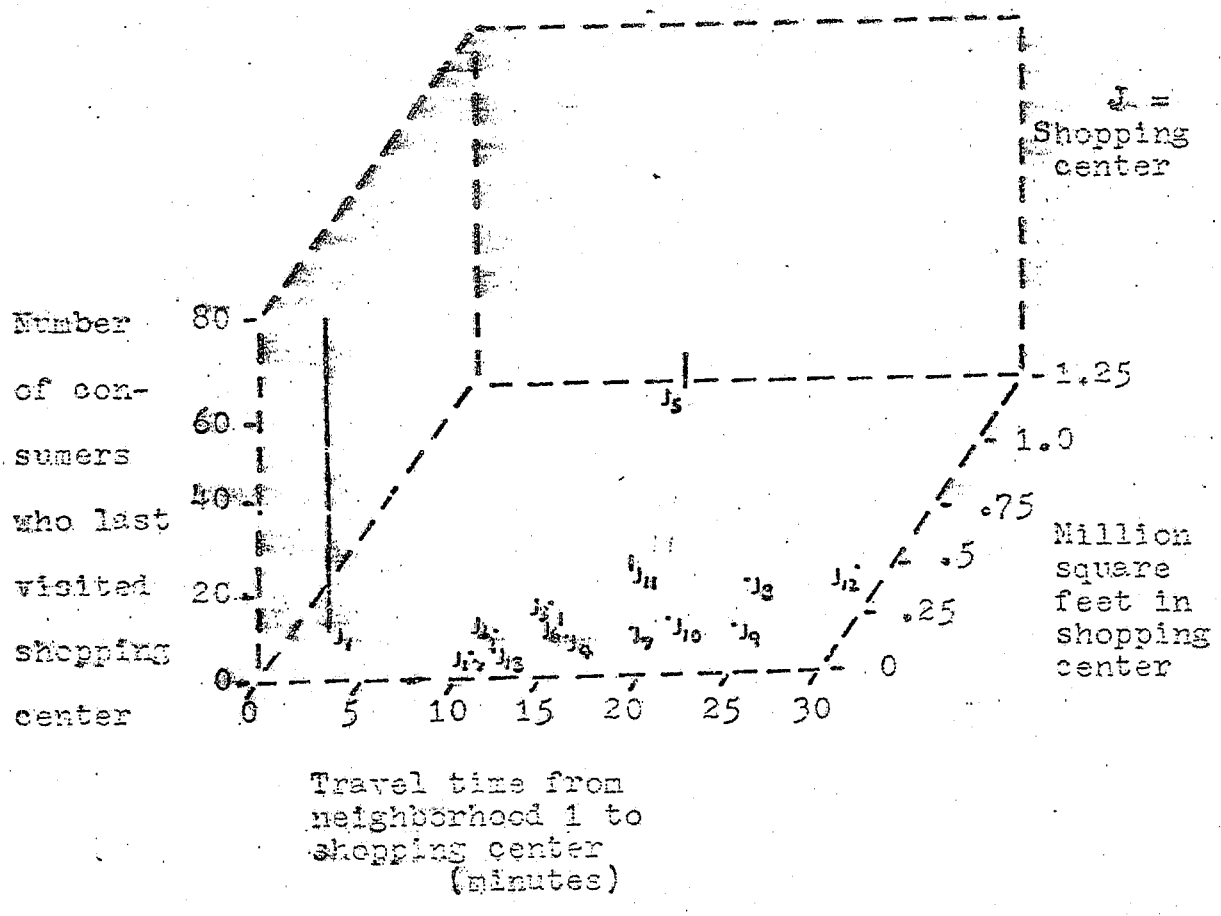
is greater than unity, the plane takes on a shape similar to that in figure 2.3, where its slope always exceeds 45°. If λ is less than unity, the slope of the plane never exceeds 45°. Clearly, if λ can be successively varied until that plane is found (by the "sum of least squares fit" method) which most closely fits the plotted observed behavior (see figure 2.4), the "predictive" ability of the model is inevitably increased.

The apparent accuracy of the model's predictions is, however, in doubt, especially in view of the few actual interactions and the extreme values related to each (see table 2.2). As has been demonstrated, the value of λ determines the angle of slope of the plane along the log T_{ij} axis. The reason for the accuracy in predicting the number of T_1 interactions (see table 2.2) is that an isolated large value such as T_1 exerts a strong influence on the path of the plane being fitted to the data. The existence of four very small frequencies results in the plane passing close to, but not through, these values by the "sum of least squares fit" method. Inevitably, the close proximity on the graph between several of the patronized and non-patronized J's (see figure 2.4) results in some predictions of interaction with centers which had not been visited.

To fit a plane surface, by varying the value of λ, to a set of interaction frequencies with such remarkably

FIGURE 2.4[8]

The Spatial Interactions of Neighborhood Residents with the Selected Shopping Centers



[8] Figure derived from data in Huff's paper, op. cit.

extreme values and with so few places actually patronized (see figure 2.4), and thereby obtain accurate results, is hardly surprising. Given any set of data having only very high and very low values, the fitting of an accurate line or plane to the data, by simple or multiple regression analysis, is assured. If Huff had tested the model in a situation with a much more complex choice of alternatives and for many neighborhoods, rather than for one, the results would have more significance. The larger the number of frequencies lying between the extreme ranges of Huff's data, the greater the probability of a decline in the model's predictive power. For, in order to retain its existing accuracy the intermediate frequencies would have to occur in a narrow zone on either side of the plane. In other words, the very simple consumer preference ranking, which Huff's model postulates, would have to be replicated in a much more complex set of spatial opportunities. Even then, the criticisms made above, regarding the difficulty of interpreting results from regression analysis and its inability to make meaningful statements about spatial behavior, irrespective of consumer spatial situation (see page 19), cast doubt on the technique's appropriateness to spatial behavior in complex spatial situations. Similarly, for the same reason that regression analysis was criticized, the usefulness of Huff's measures of cardinal utility is questioned. Certainly, in themselves, they can be argued to be possible measures of

cardinal utility. But, insofar as the value of regression analysis and analogous gravity models has been doubted, the value of a technique of measuring cardinal utilities, derived from a gravity model, is equally open to question.

In the original choice of neighborhoods, Huff selected each of the three sample neighborhoods on the basis of their homogeneity of population density and household income. "It was therefore expected that lambda would be approximately the same for each neighborhood with respect to a given type of shopping trip, i.e., clothing and furniture."[9] For it can be postulated that the selection of homogeneous socio-economic areas reduces the likelihood of encountering different sets of space preferences. If each neighborhood has the same preference ranking (i.e. if space preferences are absent) then the equations describing these similar rankings are, by definition, the same. However, Huff, in fact, obtained three different values of λ in the gravity equation (see table 2.1). Whilst he demonstrated statistically that the three values could come from the same population, the limited number of degrees of freedom impose very wide confidence limits on the null hypothesis that the three observations could be drawn from the same population. Thus it is not wholly improbable that the λ value for neighborhood three is significantly different.

[9] Huff, D. L., op. cit., pp. 455 - 456.

However, even if all three lambda values were shown to be significantly different, no inference could be made regarding the existence of space preferences. For it is impossible to impute differing lambdas to the existence of different sets of preference rankings, unless the consumers compared are similarly situated in space with respect to alternative retail outlets.[10] If differently situated consumers do not behave similarly in space, these behavior patterns may as well be a function of the initial differences in situation, as of space preferences. For instance, it is very likely that the spatial behavior pattern of consumers located close to a large retail outlet will be different from that of consumers at some distance. Now, if the former are a low-income group, and the latter high-income, it is not possible to distinguish whether location or economic status, or both, are the variables correlated with different behavior patterns. In this respect, Mitchell's criticism of regression models of spatial behavior, mentioned above, is appropriate. Huff's model, therefore demonstrates the limitations and logical loopholes in this application of regression analysis. It is also demonstrably unable to test for the significance of

[10] Alternatively, the concept of indifference curves can be employed in such a way, that differently situated groups can nevertheless be compared as if they were similarly located. This argument is explained in 2.3.

space preferences, and so would not serve as a useful basis for the present study.

2.3 The Indifference Curve Model

So far, it has not been possible to attribute predictive error in models specifically to spatial or non-spatial variables, as the method used in these models has been unable to cope properly with consumers differently located in space. Error in prediction, therefore, may be simply a function of this model inadequacy and not attributable to the existence of space preferences. Indifference curve analysis of consumer spatial behavior does permit testing of the hypothesis that different space preferences exist and are revealed in consumer spatial behavior.

The concept of indifference curves has been most fully discussed in the literature of micro-economics and consumer choice. If "utility" might be used interchangeably with the term "subjective value", then "an indifference curve is, in Edgeworth's formulation, a constant-utility curve"[11]. A consumer is hypothesized to be indifferent between any combination of two commodities (which can be quantified) on that curve. In conventional economics, a greater quantity of both of the commodities will lie on a

[11] Edwards, W., "The Theory of Decision Making", Psychological Bulletin, Volume 51, Number 4, 1954, p. 384.

higher indifference curve (see figure 2.5). In terms of Rushton's model, the two "commodities" are town size and distance from household to town. The latter "commodity", however, has negative utility and so a larger quantity of the town size commodity and a smaller quantity of the distance commodity will lie on a higher curve; hence the different shape of a town size/distance indifference surface (see figure 2.6).

If an indifference curve is an expression of constant utility, then an indifference map or surface, represents a ranking of utilities, defined in this study as a ranking of alternative spatial opportunities for consumer expenditure, i.e. a space preference structure. In order to avoid unsatisfactory preference structures (rankings), it is necessary to assume both, that the consumer has complete information and that he has "a complete.....weak[12] ordering for all commodity combinations, or points in commodity space"[13]. The ranking described by an indifference surface, however, can only be measured on an ordinal scale in the present study, although it has been argued that "very little extra effort is needed to reach a

[12] A weak ordering exists when, although there is a complete ordering of all commodity combinations, a consumer is nevertheless indifferent between certain combinations.

[13] Edwards, W., op. cit., p. 385

FIGURE 2.5

A Hypothetical Indifference Map

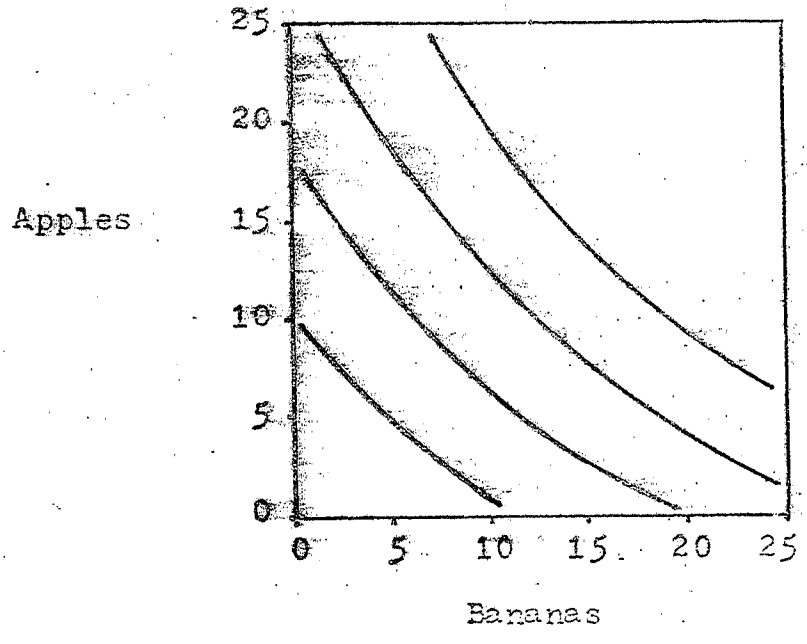
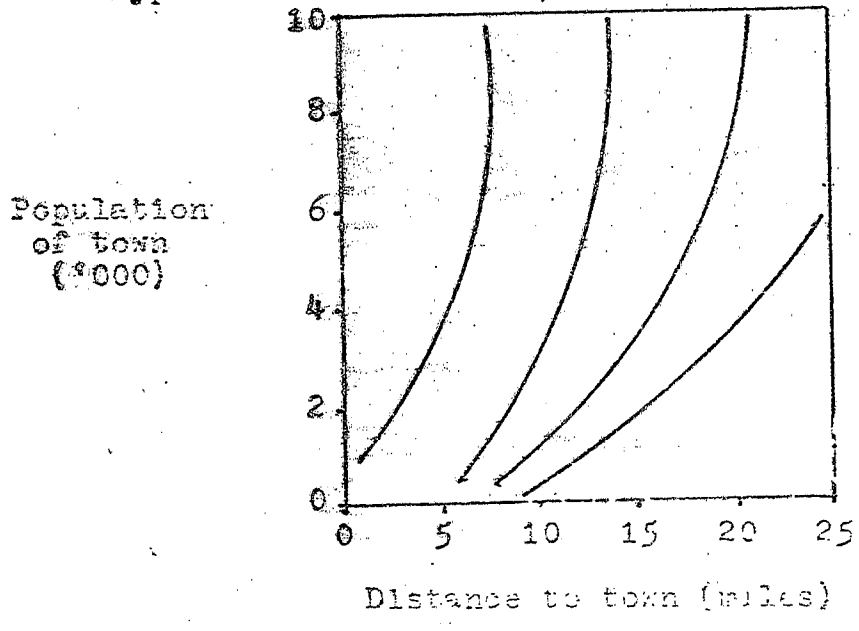


FIGURE 2.6

A Hypothetical Town Size/Distance Indifference Map



numerical [or cardinal] utility"[14]. In other words, the surface can only predict whether a consumer prefers combination A to combination B, but not the amount by which A is preferred to B. Pareto has argued[15] that this accords with reality, in that people can tell whether they prefer to be in state A or state B, but cannot tell how much more they prefer one state to another. Similarly, it can be argued that a consumer's revealed preference ranking of alternative retail outlets is purely ordinal, unless relative dollar expenditure in each outlet is regarded as an index of cardinal utility.

In fact, Rushton makes no use of cardinal measures, since the model is concerned to predict only the maximum purchase town or first ranked town for each household. Thus, the difference between cardinal and ordinal ranking is irrelevant. In contrast, both Mitchell and Huff make use of the cardinal utilities predicted by their models. In Mitchell's, a complete cardinal ordering of towns is developed for each household, insofar as the model predicts each household's dollar expenditure in every town it is predicted to patronize. Similarly, Huff's model implies that each neighborhood has a set of cardinal utilities ascribed to

[14] von Neumann, J., and Morgenstern, O., "Theory of Games and Economic Behavior", Princeton University Press, Princeton, 1944, p. 17.

[15] Pareto, V., "Manuale di economia politica, con una introduzione ulla sciensa sociale", Milan, 1906.

the fourteen shopping centers. This is revealed in the prediction of the relative proportion of a consumer's interaction with every center (P_{ij}).

In one other respect, the indifference model differs significantly from the regression and gravity models; namely, it is more competent to compare the spatial behavior of groups differently located with respect to spatial opportunities. It will be argued that, for this reason, the present technique used to derive an indifference surface permits tests to assess the relevance of space preferences in consumer spatial behavior patterns.

As was demonstrated in Huff's model, it is impossible to determine whether the different values of lambda derived for each of the three neighborhoods is a function of different spatial situations or of non-spatial variables. Using the present technique for deriving indifference curves, in which each group's actual interactions in every spatial situation is expressed as a ratio of possible interactions, the dissimilar location of groups becomes irrelevant. Figure 2.7 tries to illustrate more clearly, why the inclusion of a ratio of actual to possible interactions, rather than a simple observation of actual behavior, overcomes the problem of dissimilar locations. In this situation, although high and low income groups are differently located with respect to each town, the fact that the number of

FIGURE 2.7

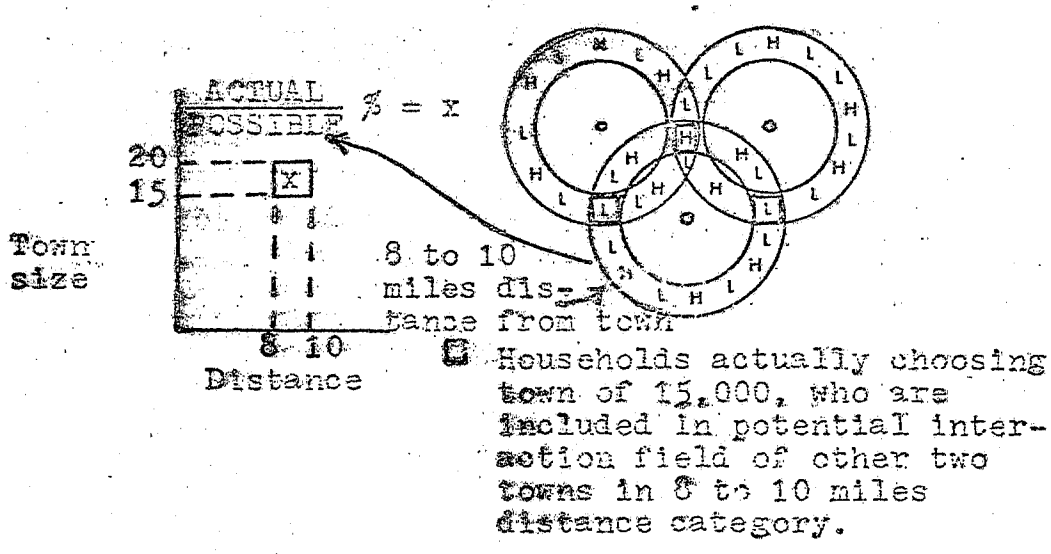


FIGURE 2.8

Goods and Services for which the Space Preferences of "Modern" Canadians and Old Order Mennonites were Compared

Good or Service

Doctor
Dentist
Bank
Appliances
Auto Repair - Harness Repair*
Food*
Clothing (and yard goods)*
Shoes*

* Significant difference between the spatial patronage patterns of the two groups.

actual interactions of each sub-group (H and L) is expressed as a ratio of that group's possible interactions with a town of given size at a particular distance, cancels out the effect of dissimilar spatial situations. Furthermore, each household, which actually interacts with a town in one spatial situation, is also included in the calculation of possible interactions with all other alternative spatial situations open to that household. Therefore, it is possible to hypothesize that any significant differences in the indifference surfaces of groups and hence, in their space preferences are not attributable to possible dissimilar locations, but to some other variable(s). In consequence, it is possible to make use of such a model in a test to assess the significance of space preferences.

The fact that the indifference surface derived in the study, explains over 65% of spatial choices, suggests that it is a close approximation to the mental rankings of the spatial alternatives of over 65% of the sample. Conversely, the incorrect explanation of the choices of the other 35% indicates that the surface is an insufficient approximation of their space preferences, to predict their spatial choices. One possible interpretation, pursued in this study, is that there is more than one space preference structure for all households in the sample.

2.4 Trade Area Studies

One other group of works concerned with consumer

spatial behavior, are trade area studies. Although not specifically structured to elucidate hypothesized regularities of spatial behavior, they commonly assume, and sometimes claim to have proven, the Central Place Theory hypothesis of interaction with the nearest place offering the good desired. A trade area study conducted by Berry, Barnum and Tennant[16] in southwestern Iowa demonstrates, according to Berry, that,

> "....the farmers make the same clear choice [in patronizing central places for goods and services]. There is only a little interdigitation 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"[17].

No exact quantities are provided to reinforce the argument, although desire-line maps are produced. The study implicitly maintains that consumer spatial behavior is a function of distance alone, and so discounts the possibility of space preferences. A test of the accuracy of such a study is to be found in a study of a sample of the dispersed Iowa population[18]. The hypothesis tested is that a consumer

[16] Berry, B. J. L., Barnum, H. G., and Tennant, R. J., op. cit.

[17] Berry, B. J. L., "Geography of Market Centers and Retail Distribution", Prentice-Hall, Englewood Cliffs, N. J., 1967, p. 16.

[18] Rushton, G., op. cit., p. 16.

will make his maximum expenditure on groceries in the nearest town to him. The definition of "nearest town" is successively varied to include only towns above a given population.

Table 2.3

Results of Test of Nearest Town Hypothesis with Various Town 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, G., op. cit., p. 16.

The trade area study is similarly "based upon responses as to where consumers purchased most of a given commodity"[19]. However, the lack of quantitative evidence to support the conclusions of the study, together with discrepancies

[19] Berry, B. J. L., op. cit., p. 23.

between the evidence of some desire-line maps and stated findings, cast doubt on the validity of the conclusions reached.

2.5 Space Preference Studies

The models discussed so far can be classified as attempting to explain spatial behavior as a function of variables related to consumer spatial situation. All have limited predictive ability and, despite methodological weaknesses in some, they lend weight to the notion that the addition of non-spatial consumer variables would improve their powers of prediction.

The writer knows of only two studies which make specific tests for the existence of space preferences[20]. In Murdie's work the spatial expenditure patterns of Old Order Mennonites and "modern" Canadians in Waterloo County, Ontario were compared. The specific behavioral characteristic measured was "distance traveled to the first choice center" where "first choice center is the place where a good is most frequently obtained"[21]. The findings showed a similarity in spatial behavior with reference to some goods and services, but marked differences in others (see figure 2.8). "Differences in space preference are an expression of

[20] Murdie, R. A., "Cultural Differences in Consumer Travel", Economic Geography, Volume 41, number 3, July, 1965, pp. 211 - 233. Rushton, G., op. cit., Chapters 4 and 5.

[21] Murdie, R. A., op. cit., p. 215.

between-group variations in the nature of transport technology [the Mennonites travel by horse and buggy] and consumer needs. These differences are also analogous to variations in space preference between most of North America today and the North America of the pre-automobile era"[22]. However, the knowledge of such space preferences was not used to improve the explanatory power of any model. The work is of value in demonstrating a relationship between spatial behavior and a consciously held set of beliefs. Thus space preferences are revealed where a conscious effort is made to behave spatially in a pattern markedly different from the norm.

Rushton analyzed several indices of consumer spatial behavior by comparing the behavior patterns of different groups, defined in terms of social and economic characteristics. Having sub-divided the sample population according to some 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 traveled 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. Whilst the study provides interesting information regarding group differences in behavior, the

[22] Murdie, R. A., op. cit., p. 232.

usefulness of the findings in improving the original model is in doubt. It is arguable that the primary concern of the geographer in analyzing consumer spatial choice behavior is to elucidate consumer preference rankings of alternative spatial opportunities. Such knowledge in the form of a statistical law, for instance, would enable further meaningful generalizations to be made and would be a manifestly important advancement of knowledge. A critical feature of human spatial behavior would be understood, and on intuitive grounds, it is reasoned that the scientific ramifications of such knowledge would be considerable. However, several of the indices in Rushton's study (e.g. "largest dollar amount spent in the farthest grocery purchase town"[23]) would seem not to be able to contribute to an understanding of how space preferences affect consumer preference rankings. Thus the knowledge gained from the tests made, though not inconsequential, is somewhat peripheral to the main focus of concern, namely consumer preference rankings.

Furthermore, the space preference tests made cannot be related to the main concern of the study centered around the development of an indifference surface or surfaces adequately explaining spatial behavior. It is impossible to evaluate, whether the preferences revealed in the latter

[23] Rushton, G., op. cit., p. 66.

part of his study are significantly related to the model's incorrect explanation of 35% of behavior. If they are related, then it should be possible to show a significant concentration of a given type of consumer with a certain space preference in either the correctly or incorrectly predicted group of households.

Finally, no analysis was performed on each social and economic grouping of households to ascertain whether they were similarly located in space. Thus, without evidence to the contrary, 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 space preferences, or of both.

Thus, the task remains of establishing whether consumer space preferences can significantly affect the explanatory power of a model incorporating only variables related to the consumer's spatial situation. Significantly, however, both of the above studies do conclude that space preference is in fact a relevant dimension influencing consumer spatial behavior.

2.6 Huff's Topographical Model[24]

In this work, Huff develops a sui generis model

[24] Huff, D. L., "A Topographical Model of Consumer Space Preferences", Papers and Proceedings of the Regional Science Association, Volume 6, 1960, pp. 159 - 173.

incorporating several groups of spatial and non-spatial variables, intuitively deduced to be related to revealed consumer spatial behavior (see figure 2.9). Those variables described under the headings "behavior-space perception" and "movement imagery" can be defined as "spatial", whilst those related to the consumer's "value system" are essentially non-spatial in character. Although Huff attempts to deduce the connectivity between all variables in the model, the study cannot utilize similar intuitive deductive methods to evaluate the connectivity between each variable and overt behavior, and more fundamentally, the relative degree of connectivity between P_4, P_{15}, P_{21} and overt behavior. The aforementioned "spatial" studies have demonstrated the relationship between variables, which could be subsumed under the P_{15} and P_{21} headings, and overt behavior. The present study concentrates on the relationship between factors thought to be related to the consumer's value system, and his overt behavior.

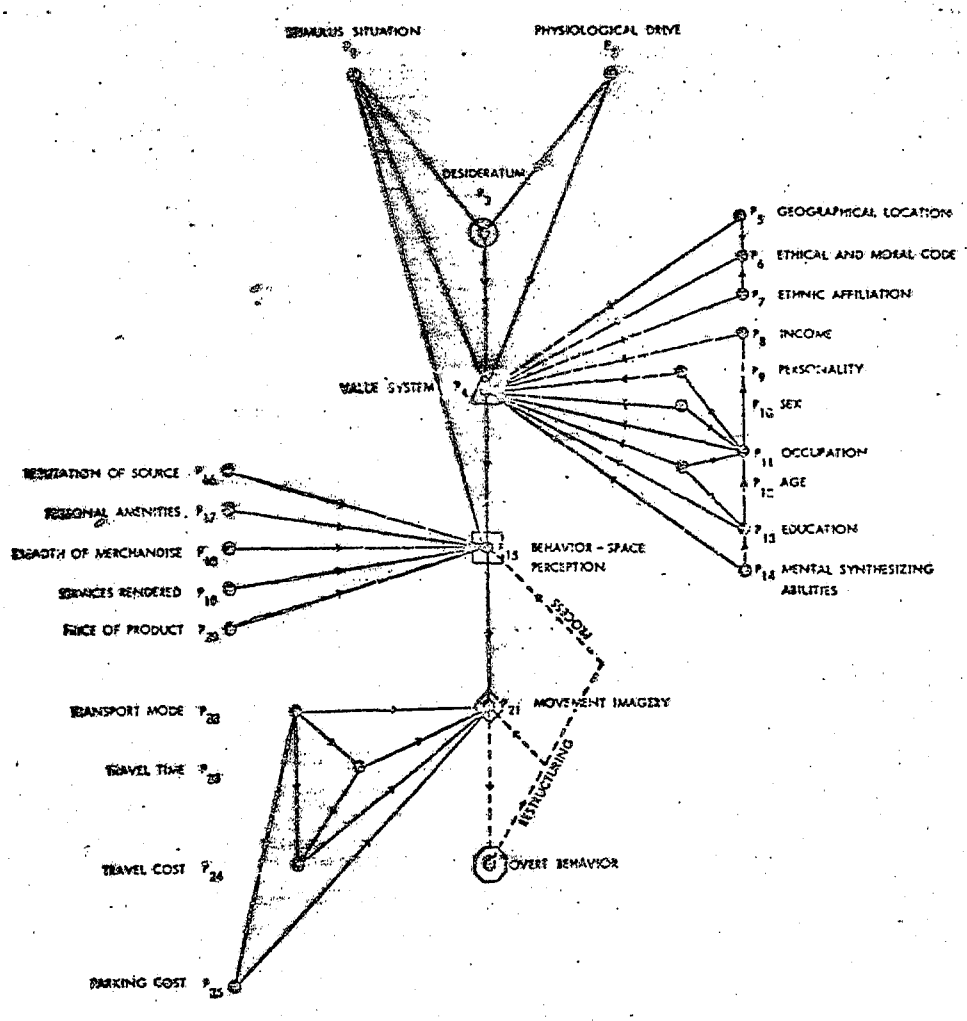
2.7 Trip Frequency Studies

There does exist a body of geographical literature in which the relationship between consumer spatial behavior and socioeconomic variables is studied. Marble, in a comprehensive review of this literature[25], finds that several authors have been able to relate one aspect of

[25] Marble, D. F., op. cit.

FIGURE 2.9

Basic Interactions of the Model



Source: Huff, D. L., "Consumer Space Preferences", op. cit., p. 165.

spatial behavior to the socioeconomic characteristics of individuals and households[26]. In every case, however, this pattern of spatial behavior is defined as "trip frequency". The regularity observed, therefore, is not so much spatial as temporal; the dimension being emphasized, more related to time than space. Amongst the variables found to be significant explainers of variations in trip frequency are, age of respondent, occupational structure of household, wealth (in relation to non-work trips)[27], family size, and automobile ownership[28, 29]. Findings are not consistent with each other in some cases, although this may be a result

[26] Gardner, J., "A Study of Neighborhood Travel Habits in Baltimore, Maryland", unpublished M. A. thesis, Cornell University, 1949.

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

Mertz, W. L., and Hamner, L. B., "A Study of Factors Related to Urban Travel", Public Roads, Volume 29, April, 1957, pp. 208 - 212.

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.

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

[27] Gardner, J., op. cit.

[28] Hamburg, J. R., op. cit.

[29] Mertz, W. L., and Hamner, L. B., op. cit.

of differences in research design or purpose. Nevertheless, these studies do not seek to say anything meaningful about spatial interactions, per se, but rather about a temporal facet of such interactions.

2.8 Conclusion

A gap remains, therefore, in the body of geographical literature relating to consumer spatial behavior; namely, the hypothesized existence of space preferences has never been rigorously tested, despite frequent mention in the literature. The above review of relevant literature has sought to point up the various methods used to analyze consumer spatial behavior, as well as their methodological differences and where applicable, their methodological limitations. It has also been concerned to suggest that part of their weakness is attributable to the absence of non-spatial consumer variables in the models. The only works including such variables in their analyses are not directed to a study of consumer spatial behavior, per se, and so do not provide an answer to the existing problem. This study endeavors to provide a partial answer, at least, to the question of the significance of space preference in consumer spatial behavior.

CHAPTER 3

The Test of the Space Preference Hypothesis

3.1 The Hypothesis and Method of Testing

In previous discussion, space preference was defined as a consumer's mental ranking of all conceivable spatial locations (1.2). In addition, an indifference surface has been explained as an expression of the consumer's space preference structure (2.3). On the basis of this definition of the indifference surface, it was suggested that the 35% incorrect prediction of consumer spatial choice in Rushton's model (2.3), could be interpreted to mean that more than one space preference structure is present in the sample population. This line of reasoning led to the goal of the present analysis which is to test for the existence of space preferences in this population, in relation to specific social and economic variables[1].

However, it is feasible that there is, in fact, only one space preference structure and that the indifference surface is not able to fully replicate that ranking, on account of some undetermined factor other than space preferences. If this factor were to account completely for the model's inaccuracy, it is

[1] Psychological consumer characteristics are not discussed in this study.

axiomatic that no significant difference in space preferences could be established in the following analysis. If however, the other factor were to account for only part of the error, it is possible that some indication of space preferences would be given by the analysis. The question as to how much the results of the model's predictions reveal about differences in consumer space preferences and how much about the suitability of the model as an index of these preferences, is tackled in this and the following chapter.

If indeed the apparent difference in space preferences of the correctly and incorrectly predicted groups, indicated by the model's partial inaccuracy, is related to socioeconomic traits of consumers, then it is reasonable to suppose that the two groups will have different socioeconomic compositions. The goal of the following analysis is to establish which social or economic variables these may be, and the form of their relationship to differing space preferences. For each of fifteen variables studied, the null hypothesis tested is that the statistical distribution of the variable in the explained and unexplained groups of households is not significantly different. In other words, the null hypothesis is that the two groups could be regarded as being randomly drawn samples from the same population, with regard to that variable. If, in fact, the statistical distribution of values of a variable is

significantly different between the two groups, a relationship can be inferred between the difference in values of the variable, and the apparent difference in the space preferences of the two groups.

To ensure that the above inference is not a false one, as well as to determine the form of the relationship between the variable and consumer space preferences, a further analysis is required. For, if the relationship is a significant one, there should be some differences in the space preferences of consumers with different "scores" on that variable. Assuming that households characterized by higher values of a given variable have a different space preference structure from households with lower values, the sample population is arbitrarily divided into two groups[2], the one with high values and the other with low values of the variable. Indifference surfaces can be derived for each group, using only its own actual spatial choices and possible interactions to obtain the surface. As each surface represents a group's space preference structure, the rankings can be compared to establish what form, if any, the difference in space preferences takes.

The following are generalized examples of how

[2] For the purposes of the analysis to be described, the problems of sampling error and of an insufficient sample size would be greatly increased, if the population was further sub-divided.

inferences can be made about consumer space preferences, by visual comparison of the indifference curves of more than one group (see also page 55).

DIAGRAM 3.1

i) Group 1: B is more attractive than A
Group 2: A is more attractive than B, i.e. larger town at greater distance is preferred to smaller town at lesser distance

ii) Group 1: B is preferred to A, is preferred to C
Group 2: C is preferred to A, is preferred to B

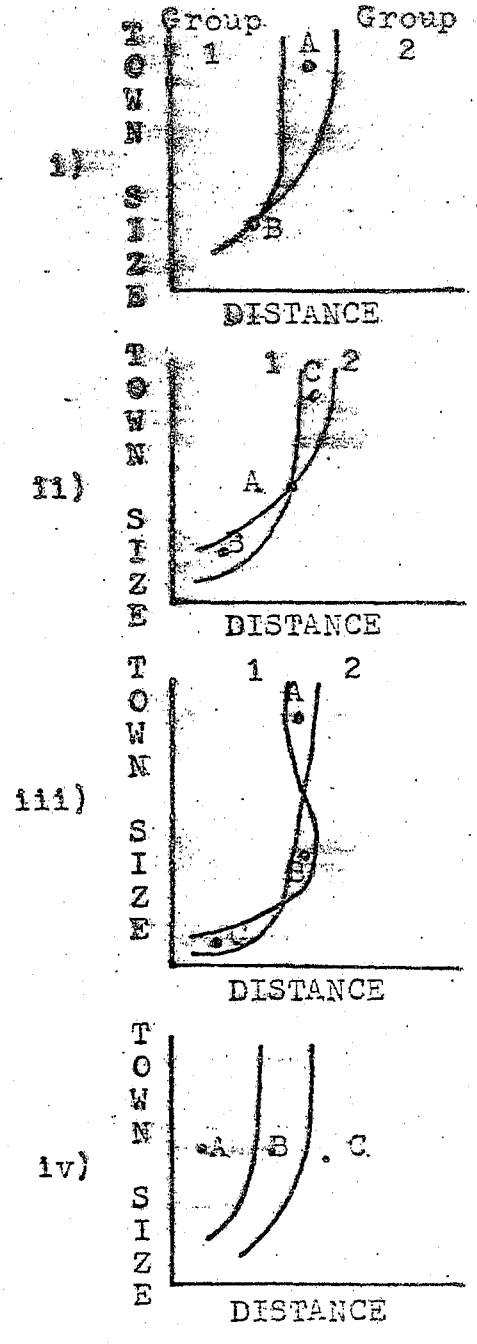
iii) Group 1: B is preferred to A and C
Group 2: A and C are preferred to B

iv) Where the curves take the same shape there is no difference in the preference rankings of the two groups, i.e. both prefer A to B and B to C.

3.2 The Test for Space Preferences

Whereas an indifference surface derived from the behavior of a population with only one space preference structure can be expected to accurately predict their spatial choices (see Appendix 2), it is probable that a surface derived from the behavior of a population with more than one space preference structure will have more limited predictive power. The surface derived in the latter case would be an approximation to the mean of the several space preferences and would therefore be less able to predict the spatial choice of an individual having any one of the space preferences. Assuming, therefore, that the incorrect explanation of the spatial choices of 35% of the population is in part a function of differing space preference structures

DIAGRAM 3.1


i)
Group 1
Group 2
A
B
TOWN SIZE
DISTANCE
ii)
1
2
C
A
B
TOWN SIZE
DISTANCE
iii)
1
2
A
B
C
TOWN SIZE
DISTANCE
iv)
A
B
C
TOWN SIZE
DISTANCE

being incorporated in the original model's indifference surface, it is reasonable to expect that a surface separately derived from the spatial choices and possible interactions of only the incorrectly explained group would more closely approximate the space preference structure of this group. In fact, this is borne out in the 30% correct prediction of the "unexplained"[3] group's behavior from its own indifference surface (figure 3.1). The inference to be drawn here is that the indifference surface in figure 3.1 is a closer approximation to the "unexplained" group's space preference structure(s) than the indifference surface derived for the total sample (figure 3.2), which predicted the spatial choice of none of that group. The town attractiveness indices from which these surfaces were constructed are shown in table A 3.1.

A possible explanation of the fact that only 30% of the "unexplained" group's spatial choices were correctly predicted, is that this group is in fact characterized by two or more space preference structures - the same hypothesis as is presently being tested for the total population. Clearly, one could argue for the derivation of a further

[3] The terms "explained" and "unexplained" are used throughout the study to refer to the predictions of consumer spatial choice in the original model. This indifference surface was derived from the spatial choices and possible interactions of the total sample population.

FIGURE 3.1

Indifference Surface for Choice
of Maximum Grocery Purchase Town
by Indirectly Predicted Households

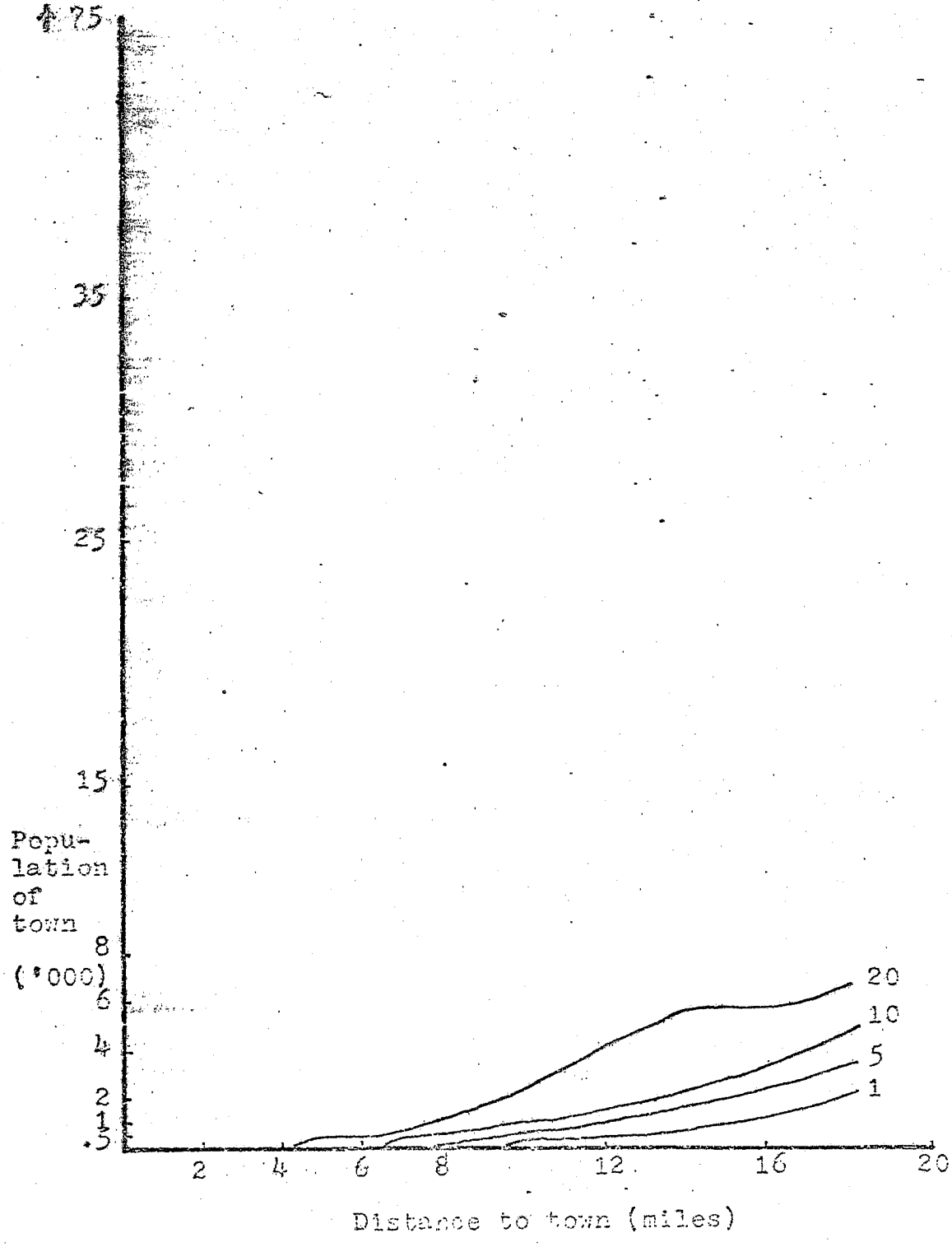
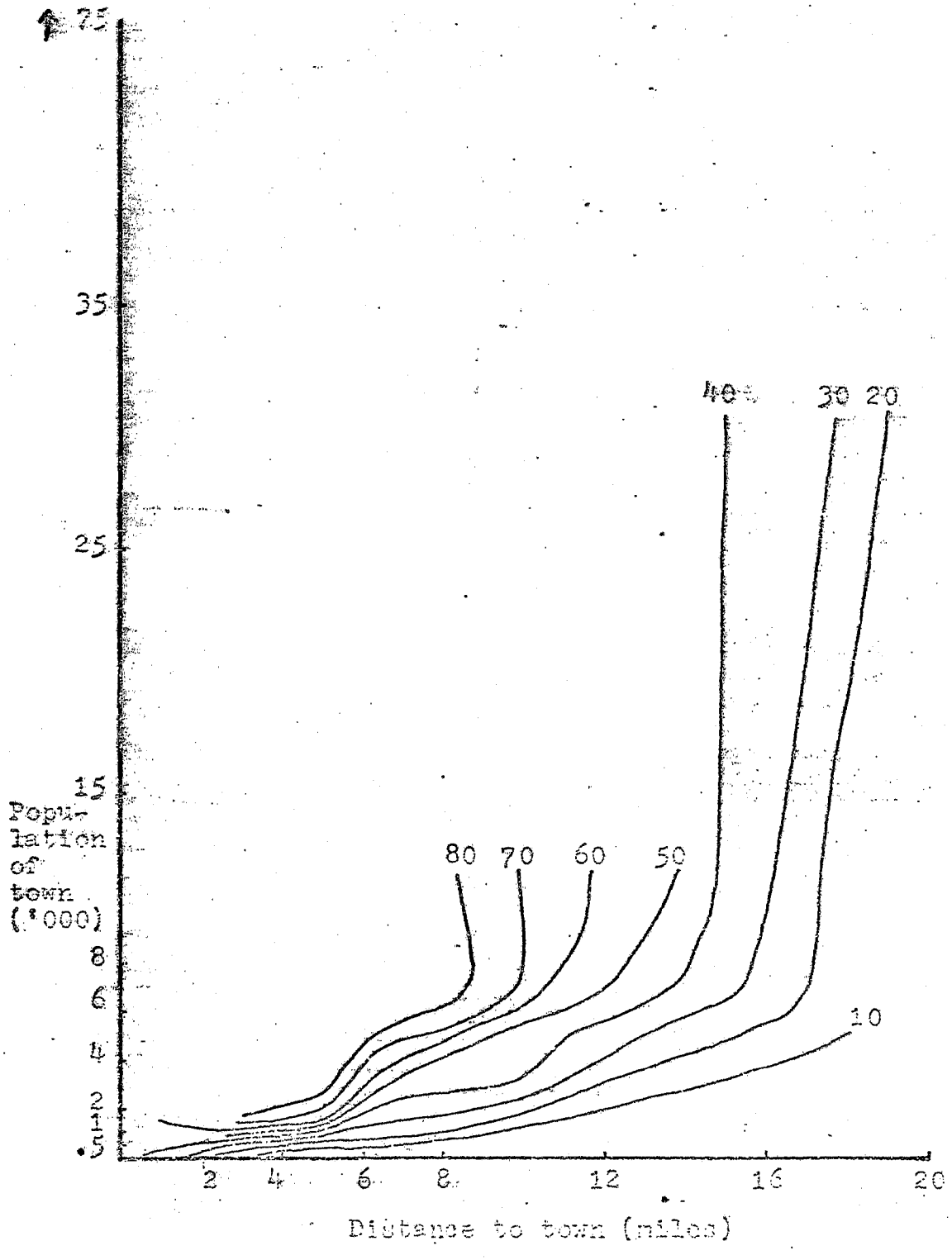


FIGURE 3.2

Indifference Surface for Choice of Maximum Grocery Purchase Town by All Households



indifference surface for that part of the group, whose behavior is not explained by the new surface. The reason for not pursuing this line of enquiry is essentially that the sample size from which one further indifference surface would be derived, is too small (110 households). With this size of sample, sampling error becomes large in the calculation of town attractiveness indices from the ratio of actual to possible interactions. The resultant surface could not, therefore, be regarded as representative of the mean of the space preference structures of that group. Its utility in explaining spatial choices would, as a consequence, be greatly diminished. Furthermore, in view of the fact that there is an alternative explanation of the model's inaccuracy (discussed in Chapter 4), it is felt that to pursue the above line of enquiry with such a small sample would be unwarranted.

Although the 30% correct prediction of the behavior of the "unexplained" households, suggests the existence of differing space preferences within the population, it gives no clue as to whether these differences are randomly distributed or related to some consumer characteristic(s). Hence the analysis described in 3.1 is employed. The hypothesis tested is that there is a statistically significant difference in the distribution of certain consumer characteristics in the explained and unexplained groups, and that those characteristics found to differ are related

to the hypothesized difference in the space preference structures of the two groups. The fifteen social and economic characteristics of households, whose statistical distribution within each group was compared, are described in table 3.1.

TABLE 3.1

Social and economic characteristics of households

1. The number of persons in the household.

2. The number of persons in the household, ten years of age or less.

3. The number of persons in the household, between eleven and twenty years old.

4. The composition of the household categorized under the following headings:-

   i) family consists only of adults under the age of 40,
   ii) family consists only of adults, 40 years of age or more,
   iii) family consists of adults, and children under 11 years of age,
   iv) family consists of adults, and children in the 11 to 20 age group,
   v) family consists of adults, and children of both the above age groups.

5. The age of the homemaker.

6. Years of education of the homemaker.

7. The age of the farm operator.

8. Years of education of the farm operator.

9. The number of members of the household working off the farm.

10. Number of years the household has lived in this house.

11. Number of years the household has operated this farm.

TABLE 3.1 (Cont'd.)

12. The total household income.

13. The net farm income.

14. Farm acreage.

15. The household's total grocery bill.

For all, except variable 4, parametric tests could be used, specifically the F test which provides a means of determining whether the two groups, described in terms of a particular variable, could be regarded as being randomly drawn from the same population; and the t test which enables statements to be made regarding the differences between the mean of a variable in the two groups[4]. Parametric tests can be used where the numerical value ascribed to a variable, such as income or years of education, has significance in terms of a numerical relationship between different observations of the same variable. Thus, a numerically significant relationship is expressed in the statement that household A earns $10,000 per annum and household B earns $5,000 per annum. However, where the "composition of a household" is denoted by a numeral with only nominal meaning, a

[4] The t-tests and F-tests were performed using a Fortran II computer program developed by Drs. Snider and Norton at the University of Iowa. For a description of the computational proceedings in both of these tests, the reader is referred to any standard text of statistical analysis, such as, Walker, H. M. and Lev, J., "Statistical Inference", New York, 1953.

parametric test is inappropriate[5], and in this case the χ^2 test for two independent samples was used. This provides a means of determining whether the two groups differ significantly with respect to the relative frequency with which group members fall into several catagories.

The χ^2 test, together with the t-test and F-test, have confidence limits placed on them, so that if the computed value of t, F or χ^2 falls within the limits, the null hypothesis of no significant difference in a variable's statistical distribution in the two groups is accepted. If a 90% confidence level ($\alpha = 0.1$) is placed on the hypothesis, this implies a 90% certainty of not rejecting a true null hypothesis. As increasing certainty of not rejecting a true null hypothesis is sought, so the confidence limits must be extended, and hence the probability of accepting a false

[5] Siegel makes this point clearly in "Nonparametric Statistics for the Behavioral Sciences", McGraw-Hill, New York, 1956, on page 3 where he says,

> "In the computation of parametric statistics, we add, divide, and multiply the scores from the samples. When these arithmetic processes are used on scores which are not truly numerical, they naturally introduce distortions in those data and thus throw in doubt any conclusions from the test. Thus it is permissable to use the parametric techniques only with scores which are truly numerical. Many nonparametric tests, on the other hand, focus on the order or ranking of the scores, not on their "numerical" values, and other nonparametric techniques are useful with data for which even ordering is impossible (i.e. with classificatory data)."

null hypothesis necessarily increases. The 90% confidence level is the lowest of three used in this analysis, since the next lowest level in common use, the 80% level, implies a one in five chance of accepting a false hypothesis. The results of the tests are shown in table 3.2. The only variable found to be significantly different between the explained and unexplained groups is "the number of years the household has lived in this house". The mean for the explained group is 15.8 years, in contrast to 20.1 years for the unexplained group. This difference was significant at the 90%, 95% and 99% confidence levels. Thus it is established that two groups with apparently dissimilar space preferences, also have significantly different statistical distributions of a given variable.

The space preference hypothesis is that the difference between the explained and unexplained groups in scores on the significant variable and their apparent dissimilarity in space preferences are functionally related. Inevitably, a linear relationship is assumed - the product both of limited knowledge of human spatial behavior, and of the absence of more complex mathematical functions. On such an assumption, it is hypothesized that households characterized by a shorter period of residence in their present house will reveal different space preferences from those of longer residence. If this were true, the indifference surfaces, derived separately from the spatial choices and possible

TABLE 3.2

Results of t and F Tests on Socioeconomic Variables

Variable	t Value	Degrees of Freedom	Confidence Limits $\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$
1. Number of persons in household	1.597	449	+1.655 -1.655	+1.970 -1.970	+2.600 -2.600
2. Number of persons in household, 10 years of age or less	0.599	449	+1.655 -1.655	+1.970 -1.970	+2.600 -2.600
3. Number of persons in household, between 11 and 20 yrs. of age	1.479	449	+1.655 -1.655	+1.970 -1.970	+2.600 -2.600
4. Composition of household (χ^2)					
5. Age of the homemaker	0.537	438	+1.655 -1.655	+1.970 -1.970	+2.600 -2.600
6. Years of education of the homemaker	0.577	449	+1.655 -1.655	+1.970 -1.970	+2.600 -2.600
7. Age of the farm operator	0.051	449	+1.655 -1.655	+1.970 -1.970	+2.600 -2.600
8. Years of education of the farm operator	0.900	449	+1.655 -1.655	+1.970 -1.970	+2.600 -2.600
9. Number of members of household working off the farm	0.782	449	+1.655 -1.655	+1.970 -1.970	+2.600 -2.600

TABLE 3.2 (Cont'd.)

Variable	t Value	Degrees of Freedom	Confidence Limits $\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$
10. Years living in this house	2.865	438	+1.655 -1.655	+1.970 -1.970	+2.600 -2.600
11. Years operating this farm	0.264	412	+1.654 -1.654	+1.969 -1.969	+2.598 -2.598
12. Total household income	1.497	438	+1.655 -1.655	+1.970 -1.970	+2.600 -2.600
13. Net farm income	0.832	393	+1.654 -1.654	+1.969 -1.969	+2.598 -2.598
14. Farm acreage	0.142	449	+1.655 -1.655	+1.970 -1.970	+2.600 -2.600
15. Household's total grocery bill	0.035	438	+1.655 -1.655	+1.970 -1.970	+2.600 -2.600

TABLE 3.2 (Cont'd.)

Results of t and F Tests on Socioeconomic Variables

Variable	F Value	Degrees of Freedom	Upper Confidence Limits $\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$	Null Hypothesis Accepted
1.	2.550	1;449	2.73	3.88	6.73	YES
2.	0.359	1;449	2.73	3.88	6.73	YES
3.	2.189	1;449	2.73	3.88	6.73	YES
4.	(α^2) 2.34	4	7.78	9.49	13.28	YES
5.	0.289	1;438	2.73	3.88	6.73	YES
6.	0.332	1;449	2.73	3.88	6.73	YES
7.	0.003	1;449	2.73	3.88	6.73	YES
8.	0.809	1;449	2.73	3.88	6.73	YES
9.	0.611	1;449	2.73	3.88	6.73	YES

TABLE 3.2 (Cont'd.)

Variable	F Value	Degrees of Freedom	Upper Confidence Limits $\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$	Null Hypothesis Accepted
10.	8.157	1;438	2.73	3.88	6.73	NO - for $\alpha=0.1$, 0.05, 0.01
11.	0.070	1;412	2.73	3.88	6.72	YES
12.	2.241	1;438	2.73	3.88	6.73	YES
13.	0.693	1;393	2.73	3.88	6.72	YES
14.	0.020	1;449	2.73	3.88	6.73	YES
15.	0.001	1;438	2.73	3.88	6.73	YES

interactions of each group, would be closer fits to each group's ranking of the actual alternative situations open to them, than one surface for the entire population. As a result, it could be expected that each surface would be a more accurate explainer of the households' spatial choices, than the original indifference surface for the total sample.

On the assumption that households with a longer residence in their present house have a different space preference structure from those with shorter residence, the total sample of 458 farm households was ranked from high to low in terms of the number of years a household had lived in its present house. The sample was arbitrarily divided into two approximately equal groups of 238 and 220 households. The former had occupied their existing house from 0 to 13 years, and the latter from 14 to 75 years. With data on the actual choices and possible alternatives of the households in each group, two indifference surfaces were derived, one for each group (figures 3.3 and 3.4). The town attractiveness indices from which these surfaces were constructed, are shown in tables A 3.2.

In fact the predictive abilities of the two surfaces were not significantly different from the original indifference surface's proportion of accurate explanation of spatial choices. The original model correctly accounted for 63.6% of spatial choices, whereas the two surfaces accounted for

FIGURE 3.3

Indifference Surface for Choice of Maximum Grocery Purchase Town by Households with less than Fourteen Years Residence in Existing House

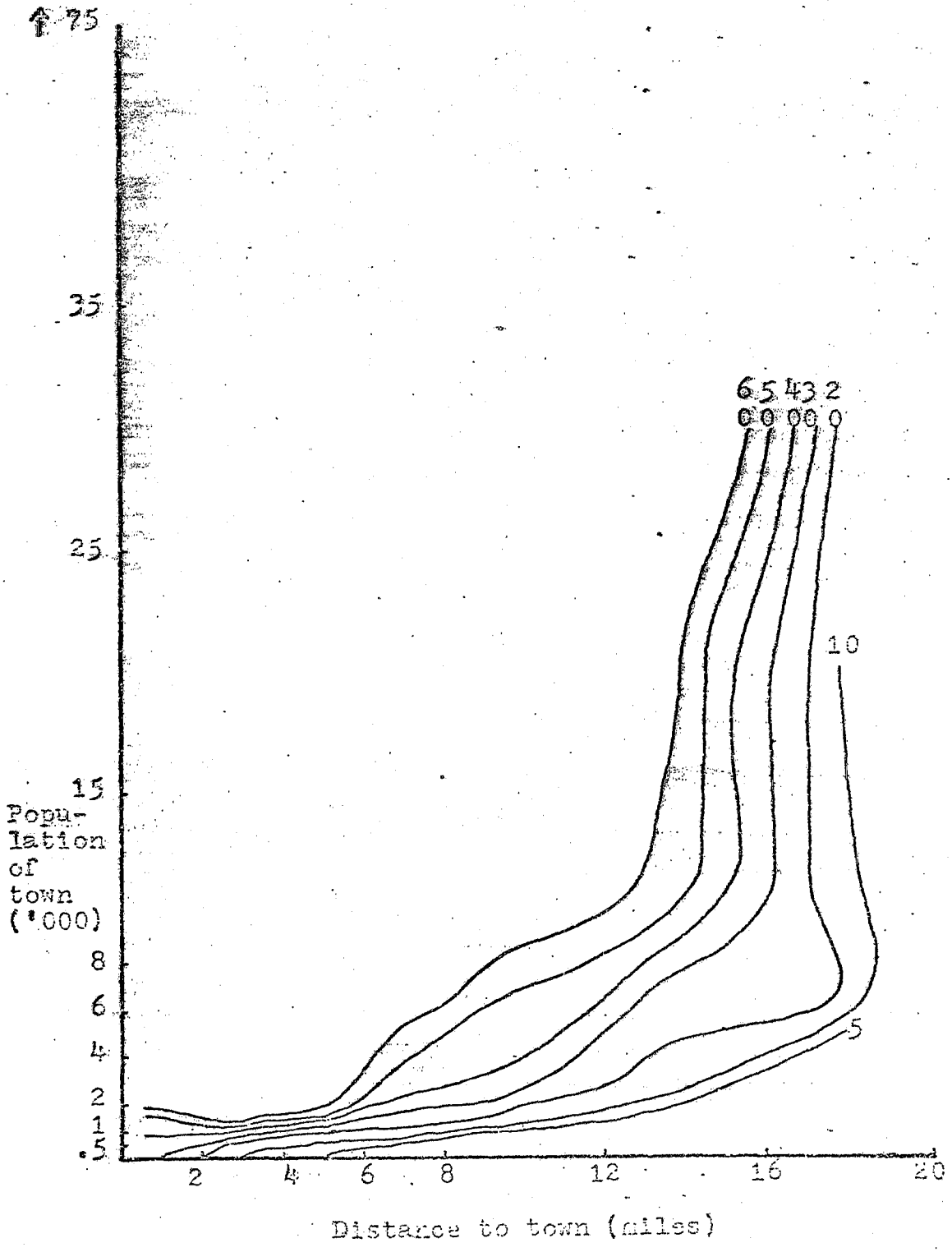
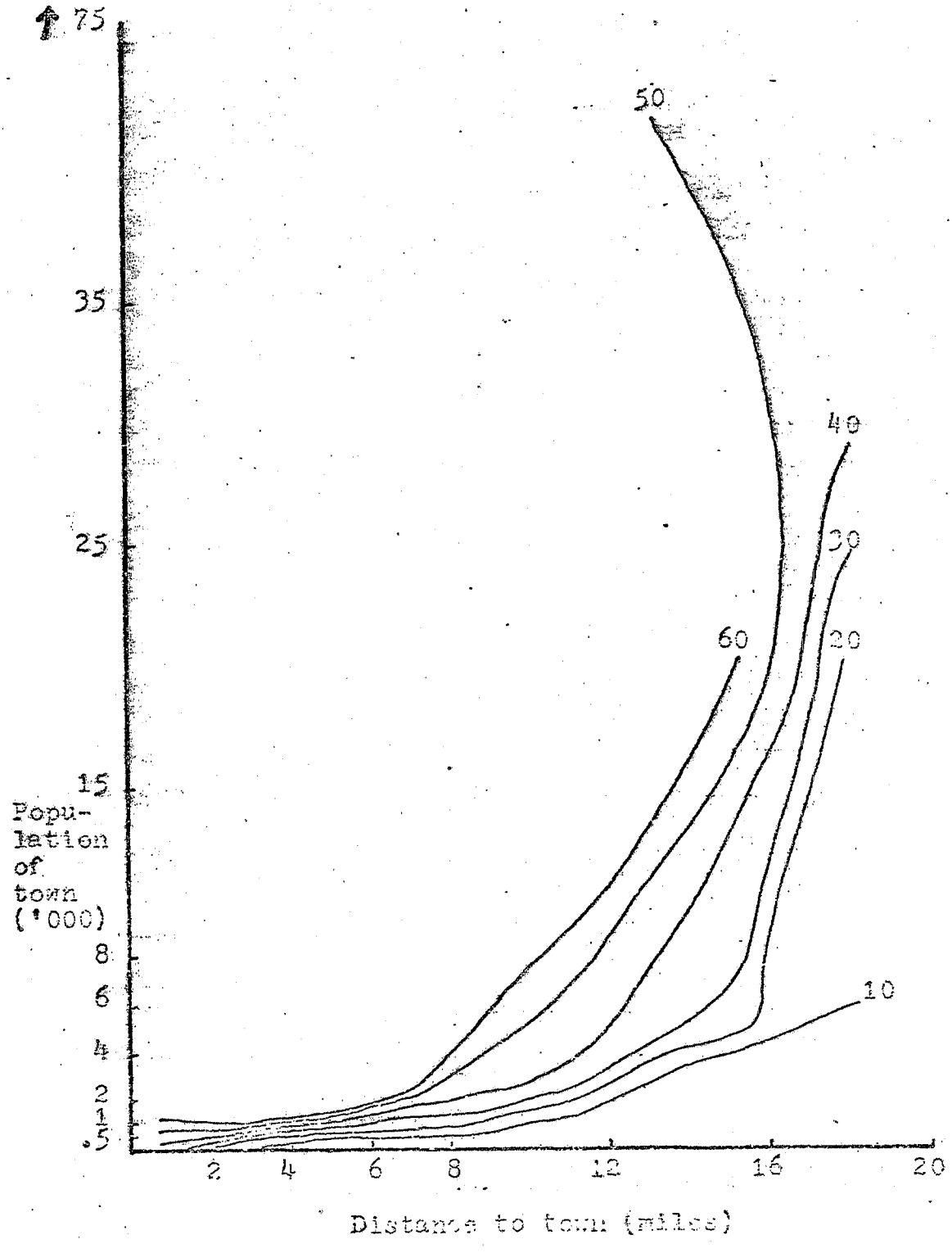


FIGURE 3.4

Indifference Surface for Choice of Maximum Grocery Purchase Town by Households with Fourteen Years Residence or More in Existing House



65.9% and 60.0% of behavior. Further analysis showed that in 83.3% of cases, the new indifference surfaces predicted the household to interact with the same town as the original model had done. Bearing in mind the distorting effect of sample size in the upper parts of each surface, where the number of actual and possible interactions is frequently very small, the surfaces are not dissimilar (see figures 3.3 and 3.4). This is particularly evident if the town attractiveness indices for towns with a population below 6,000 are compared between all three surfaces (see figures 3.2, 3.3 and 3.4).

The conclusion to be drawn from the above analysis is that the indifference surfaces derived for the two groups are not significantly different from that derived for the total population. The implication from this and the predictive similarities of the two sets of indifference surfaces, is that the new surfaces are no closer approximations to the space preference structure(s) of the population than the one surface for the total population. This in turn leads to the conclusion that households with a shorter period of residence in their existing house do not have significantly different space preferences than households with a longer residency. It is not possible to conclude, however, that a more complex division of households in terms of period of residence, also would

not reveal space preferences to be significantly related to this characteristic. Perhaps future research might iterate the above test repeatedly, each time categorizing the population differently in terms of the same variable.

The fact that only one out of fifteen socioeconomic variables had a significantly different composition in the correctly and incorrectly explained groups suggests one of two things. Either none of the other fourteen variables is related to consumer space preferences, or their possible relationship to space preferences cannot be determined by the present analysis. As was suggested above (3.1) some factor(s) other than space preferences may account for part or all of the predictive error of the original indifference curve model. If all of the error is accounted for in this way, no space preferences can be shown to exist. If, however, only part of the model's error is attributable to factors other than space preferences, an approach different from the present one may prove more useful, since the present analysis leans heavily on the socioeconomic composition of the unexplained households to indicate the socioeconomic variables possibly related to space preferences. If some members of the unexplained group are accounted for by factors other than space preferences, it becomes difficult, if not impossible, to infer which are the possibly significant variables by comparing the socioeconomic composition of the explained

and unexplained households. In this respect, it has already been shown experimentally that part of the error occurs even when predicting the spatial choices of households with one known space preference structure (appendix 2).

The following chapter seeks to determine the type of spatial choice situations under which error is more likely to occur and whether, in the light of these findings, any other factor(s) may contribute to error. If households in certain spatial situations are more likely to be wrongly predicted than in others, then a household's incorrect prediction can be related not only to the already known error factors but to its spatial choice situation. Thus, of two households each with space preferences different from that defined by the aggregate indifference surface, the household in the complex choice situation might be incorrectly predicted, whilst the choice made by the household in a simple choice situation may be correctly predicted. In other words the closeness of fit between a households actual preference ranking and that defined in the aggregate indifference surface is put to the test more in complex choice situations than in simple ones. The consequences of there being households with significantly different space preferences within the correctly predicted groups are pursued in 4.5.

CHAPTER 4

Test of the Model's Efficiency in Complex Spatial Choice Situations

4.1 The Hypothesis Tested

The alternative hypothesis to that on space preferences states that the model is unable to discriminate "realistically"[1] between alternative urban places in a complex choice situation[2]. Invariably, the indifference curve model chooses as the predicted maximum expenditure town, that one open to the consumer, with the highest town attractiveness index, no matter how small the difference between the highest and next highest index. It is argued here that the consumer's choice mechanism is not characterized by the same high degree of quantitative precision. For instance, where there is a choice between a town of 8,000 at 7 miles from the consumer and one of 9,000 at 7 miles, the model will "choose" that one with the highest town attractiveness index, assuming the index is not similar for the two towns. In reality, it seems unlikely that consumers

[1] "Realistic" is used to describe the manner in which consumers actually behave.

[2] A complex choice situation is one in which the "town attractiveness indices" ascribed to alternative spatial opportunities differ only slightly, or are similar.

perceive a noticeable difference between the two, and actual consumer behavior may be less regular in such a choice situation, than the model predicts.

The type of spatial situation in which this weakness of the model is most likely to be apparent, is where the largest town attractiveness index of any town in a given choice situation is small. It is reasoned, that the lower a household's maximum town attractiveness index, the greater the probability that the next largest will not be markedly smaller. As a corollary, the larger a household's maximum town attractiveness index, the greater the probability that its next highest index will be considerably less. In the latter situation, the model's choice is likely to reflect consumer choice, because a significant difference in town attractiveness indices indicates a significant difference between the number of actual choices in proportion to the number possible in each spatial situation. In the situation where there is only a small difference between the town attractiveness indices of two or more towns, the number of actual choices as a proportion of possible selections is not significantly different between the spatial locations. Individuals faced with a choice between these locations are unlikely to make the same invariable choice as the model.

Specifically, the hypothesis tested is that there is a significant relationship between the proportion of

interactions incorrectly attributed by a model to a given spatial situation and the value of the town attractiveness index for that situation[3]. The method of testing the hypothesis expresses the number of incorrect predictions in each spatial situation as a proportion of total predictions for that situation. Each of these proportions is then compared with the relevant town attractiveness index, to establish whether a relationship exists between the proportion of error in predicting interactions with a given spatial situation and that spatial situation's town attractiveness index (figure 4.1). It is clear from figure 4.1 that the smaller the town attractiveness index of a given spatial situation, the greater is the probability of an incorrect prediction. It appears, therefore, that whilst the model, as presently devised is able to explain consumer spatial choices in the higher ranges of the surface, the ability is weakened where the maximum purchase town predicted has a relatively low town attractiveness index. This finding is in agreement with results arrived at independently by Rushton[4].

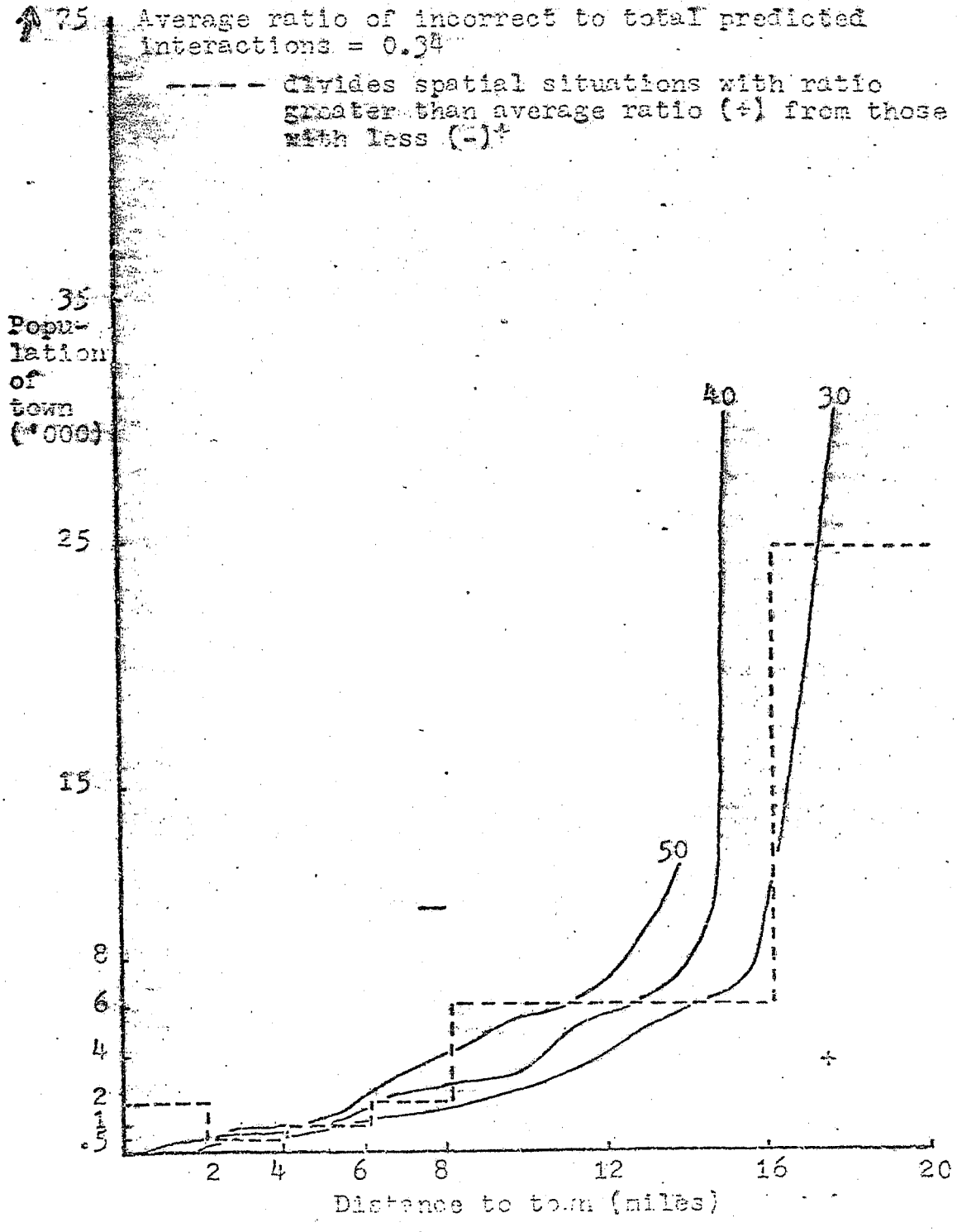
By definition, the unexplained households do not have the opportunity to interact with a town higher on the

[3] It must be borne in mind that a household is predicted to make its maximum expenditure in that spatial situation which, amongst alternatives open to it, has the highest town attractiveness index.

[4] Rushton, G., op. cit., pp. 41 - 44.

FIGURE 4.1

Zones with the Proportion of Incorrect Predictions Greater or Less Than Average*



* In figures 4.1 through 4.7, the indifference curves are those calculated for all households.

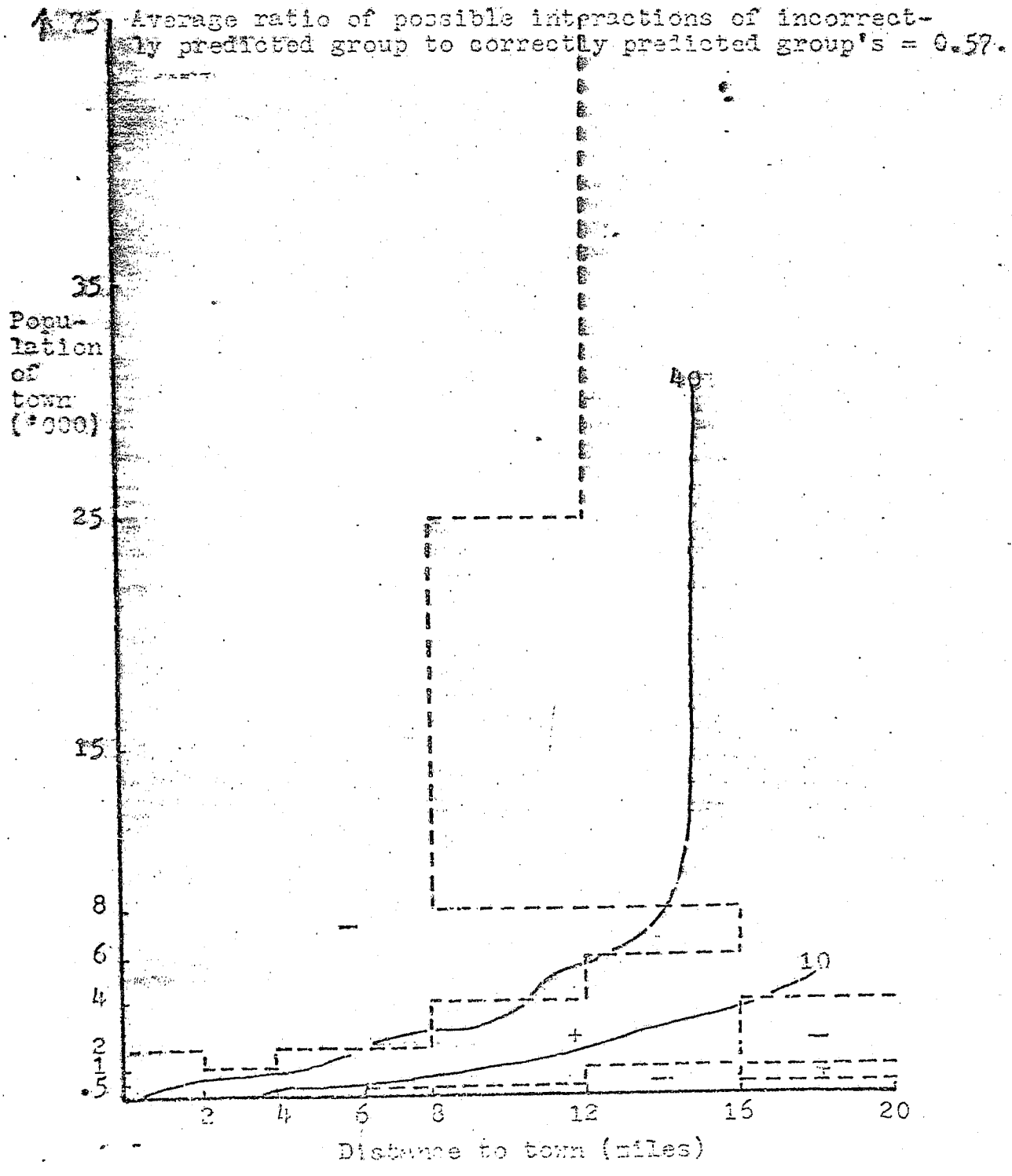
\+ This symbol has the same definition for figures 4.1 through 4.7.

surface, so that it can be deduced that these households have a more limited range of towns at various distances from which to choose. This is borne out by a comparison of each group's possible interactions (figure 4.2). There is noticeable tendency for the incorrectly predicted group's possible interactions to be more restricted to smaller and more distant towns than the correctly predicted group's. Sampling error largely accounts for the graph's form in the town size range above 8,000 (see table A 3.3 for possible interactions of each group).

Thus it can be stated that the spatial choices of households with appreciable differences in the spatial opportunities available to them are more often correctly predicted by the model, than choices where differences in alternatives are slight. The definition of differences between spatial opportunities rests on the technique of calculating the town attractiveness index. It denotes a just noticeable difference, not only between a town of 9,000 at 8 miles and one of 8,000 at 8 miles, but between any spatial situations with slightly differing attractiveness indices. Thus in the table A 3.1a) showing town attractiveness indices derived from the total population's behavior and possible interactions, there is very little difference between a town of 20,000 at 18 miles and one of 400 at 5 miles, since their town attractiveness indices are respectively 7.7 and 7.2. Invariably, where there is a choice between the

FIGURE 4.2

Zones with the Proportion of Possible Interactions of the Incorrectly Predicted Group Greater or Less Than Average



two as maximum purchase town, the model will select the former spatial situation, although, in fact that index is derived from only one actual interaction and thirteen possible. Thus the error is a function of the deterministic nature of the model, whereby that spatial situation with the highest "attractiveness", no matter how slightly different from the next highest, is invariably predicted to be the household's rank one purchase town. It would appear that the deterministic model works well in less complex spatial choice contexts, but that some form of probabilistic explanation might more closely approximate reality in the complex choice situations. For instance, if there were three alternative spatial situations with town attractiveness indices of 15, 10 and 5, each could be assigned a set of random numbers, proportionate to the size of the index; thus 1 to 15 for the first, 16 to 25 for the second and 26 to 30 for the third. The generation of a random number, say 22, would mean the model predicted the second spatial situation as the household's rank one town[5]. The point to note is that the prediction would not invariably choose the first of the three spatial situations, and that each situation would have a probability of prediction related to its attractiveness index.

[5] In principle, this suggestion is identical to Huff's probabilistic model (2.2) where the probability of a consumer at point i, interacting with spatial situation j, equals the ratio of j's utility to the sum of the utilities of all spatial situations. In the present study, utility would be assumed to be indicated by the town attractiveness index.

A further explanation of the higher proportion of incorrect predictions for smaller and/or more distant towns was tested. On the assumption that a unique ranking of spatial situations exists, it is possible that the ranking derived from the town attractiveness indices deviates from the actual ranking. Now a small difference between a computed and an actual consumer's ranking is not critical to a particular prediction of spatial choice where a person's actual alternatives are ranked far apart on his individual indifference surface. However, if the maximum purchase town and one or more alternatives were ranked close together by a consumer, a calculated indifference surface, which varied slightly from the consumer's ranking of these alternatives, might well wrongly predict his rank one town (see, for example, the differences in indifference surface in diagram 3.1). The probability of spatial alternatives being closely ranked is greatest in those spatial situations where incorrect predictions are above average[6]. As 48% of households have their rank one purchase town in one of these spatial situations, those are the spatial situations where a small deviation of the computed from a consumer's actual indifference surface would result in the greatest number of

[6] In those spatial situations where the proportion of incorrect predictions is above average (see figure 4.1) there are 6295 spatial opportunities, in contrast to 360 in all other situations. Inevitably with so many more possible interactions in the former spatial situations, the probability of closely ranked alternatives is much greater, than in the latter spatial situations with only 360 possible interactions for 448 households.

incorrect predictions by the model. Thus, if the model's computed indifference surface differed from an actual unique indifference surface, the model's higher proportion of error in these spatial situations could be attributed to the method of calculating these indices, rather than to the deterministic nature of the model in complex choice situations. To satisfy the above assumption of a unique ranking, experimental data was used, in which a known constraint is placed on the spatial behavior of the households[7]. Every household is defined to have the same set of space preferences, such that each interacts with that town having the highest $\frac{T_j}{d_{ij}^2}$ index,

where T_j = the population of the jth town,

and d_{ij} = the distance from household i to town j.

Thus every spatial situation (i.e. "town population, and distance from consumer to town" combination) has a place in the one scale of consumer preference rankings, which can be exactly described (table 4.1).

Given that the households behave in accordance with this rule, the actual and possible interactions with each spatial situation are calculated as in the original study

[7] The existing sample households and the same set of Iowa towns are used in this test. Only the behavior of the households is altered to comply with a known "rule" of spatial behavior.

TABLE 4.1

Consumers' Preference Ranking of Spatial Situations, under Constraint that Consumer Interacts with Town Having Highest $\frac{T_j}{d_{ij}^2}$

Town Population	Miles from Household to Town							
0	0	2	4	6	8	12	16	20
500	33e	49e	59e	62e	67	69	70	
1,000	24e	41e	52e	56e	64	57	68	
2,000	18e	33e	45e	49e	59e	62e	65	
4,000	12	24e	36	41e	52e	56e	61	
6,000	10	22	31	39	47	54	58	
8,000	8	18e	28	33e	45e	49e	55	
15,000	5	14	20	27	37	44	48	
25,000	3	9	15	21	30	38	41e	
35,000	2	6	13	16	26	32	40	
75,000	1	4	7	11	17	23	29	

Note: 'e' means 'of equal rank'.

(see Appendix 1). Similarly, a town attractiveness index is calculated for each spatial situation. Again, a preference ranking of spatial situations is derived from the town attractiveness indices (table 4.2), and this ranking compared with the above known space preference structure. Since each ordinal ranking can be expressed in matrix form, the second matrix is subtracted from the first, to distinguish how much the predicted ranking deviates from the known in each spatial situation (table 4.3).

It is clear that the calculated indifference surface does not replicate the known ranking. The inference, as argued above, is that those rank differences account for the model's predictive error where the consumer's rank one town and alternative spatial opportunities are closely ranked. As closely ranked alternatives are found most often in the spatial situations where possible interactions are most (see footnote 6), this accounts for the above average proportion of incorrect predictions in these spatial situations (compare figure 4.1 and table A 3.3).

Finally, the fact that there is evidence for more than one space preference structure is a further factor explaining the higher proportion of error in those spatial situations delineated in figure 4.1. The computed indifference surface does not replicate actual space preference structures both on account of the technique of

TABLE 4.2

Consumers' Preference Ranking Derived from Town Attractiveness Indices. (Constraint on Spatial Behavior same as in Table 4.1)

Town Population	Miles from Household to Town						
0	0	2	4	6	8	12	16	20
500	59	43	50	54	58	57	56
1,000	30	34	39	47	53	60e	60e
2,000	1e	19	27	42	49	55	60e
4,000	1e	24e	28	31	40	48	52
6,000	-	1e	1e	21e	36	44	51
8,000	-	1e	21e	21e	32	38	45
15,000	-	1e	1e	1e	35	41	46
25,000	-	1e	1e	-	26	24	37
35,000	-	-	1e	1e	20	29	33
75,000	-	1e	1e	1e	1e	1e	1e

Note: Rank 1 situations are those in which there are as many actual interactions as possible. Rank 60 situations are those in which there are possible interactions, but no actual ones.

TABLE 4.3

Rank Differences of $\frac{T_j}{d_{ij}^2}$ Ranking and Town Attractiveness Index Ranking*.

($\frac{T_j}{d_{ij}^2}$ Rank - Attractiveness Index Rank)

Town Population	Miles from Household to Town						
0	0	2	4	6	8	12	16	20
500	-16	+ 6	9	+ 8	+ 9	+12	+14
1,000	- 6	+ 7	+13	+ 9	+11	-	-
2,000	+17	+14	+18	+ 7	+13	+ 7	-
4,000	+11	0	+ 8	+10	+12	+ 8	+ 9
6,000	-	+21	+30	+18	+11	+10	+ 7
8,000	-	+17	+ 7	+12	+ 8	+11	+10
15,000	-	+13	+19	+26	+ 2	+ 3	+ 2
25,000	-	+ 8	+14	-	+ 4	+14	+ 4
35,000	-	-	+12	+15	+ 6	+ 3	+ 7
75,000	-	+ 3	+ 6	+10	+16	+22	+28

* No rank difference is given for those spatial situations in which there were no actual interactions, as the computed attractiveness index is zero. By definition, the model cannot predict an interaction with such a situation, as every individual clearly had a higher ranked alternative. Thus in the search for a source of the model's inability to predicted behavior correctly in certain spatial situations, it is not fruitful to include in the analysis those situations in which there were neither actual nor predicted interactions. It is axiomatic that the model could not err in these spatial situations.

calculating the expected indifference surface, and of the existence of more than one space preference structure. Thus the likelihood is greater than if a unique space preference structure existed, that the single expected indifference surface will differ from each consumer's indifference surface. In terms of the model's predictions of spatial choice, the result is that error is greatest in those cases where a consumer's rank one spatial situation is not ranked much above alternative spatial opportunities.

The conclusions to be drawn from the analyses in this section are threefold. Firstly, it is confirmed that the model's predictive error is greatest in complex choice situations. Secondly, it is argued that this circumstance is explained by the fact that whenever there is a maximum purchase town in one of these spatial situations, the probability is greater than in any other spatial situations, that there will be a closely ranked alternative. In addition, the computed indifference surface does not replicate actual consumer rankings, for the two reasons given. As a result, predictive error, due to this deviation of expected from actual rankings, is most likely where a maximum purchase town has a closely ranked alternative. The third conclusion is that some part of the model's predictive error, particularly in complex choice situations, is due to the deterministic nature of the model's predictions.

4.2 The Test Repeated in a Different Location

The finding that the model's degree of error is greatest in complex spatial choice situations prompted the re-testing of this in another location. A sample of 122 households in Grey County, Ontario was used, for which grocery purchase data was collected in 1966. The limited sample size, together with the absence of any town of over 25,000 people within 20 miles of any respondent are the major differences to be noted between this and the Iowa sample. Nevertheless, a comparison of the ratio of incorrect to total predictions of the model, with the indifference surface derived for the total population reveals a marked similarity to previous findings (compare figures 4.1 and 4.3). Despite the overall possibility of sampling error, the consistent pattern revealed by the ratios suggests the same model weakness as already described. Likewise, the range of possible interactions for the incorrectly predicted group (51% of the sample) is more restricted to spatial situations with lower town attractiveness indices (figure 4.4). In this case, the pattern is not so clear-cut as in the Iowa example, but the same generality holds true. Significantly, only 49% of behavior is correctly explained as opposed to 66% in Iowa. Part of the explanation may be in the larger proportion of complex spatial choice situations in Grey County.

Thus the analysis replicated in a different location,

FIGURE 4.3

Zones with Proportion of Incorrect Predictions Greater or Less Than Average (Grey County, Ontario).

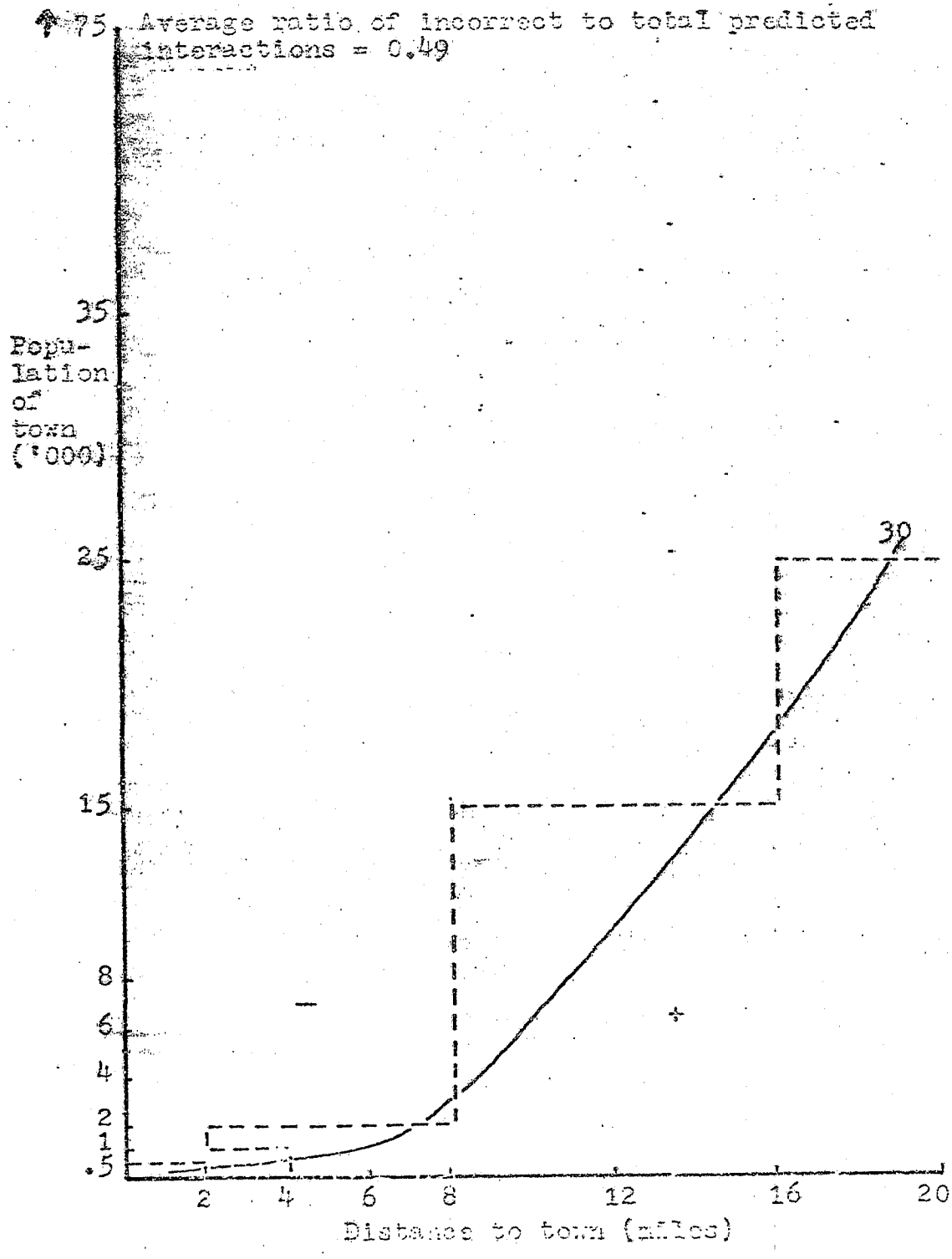
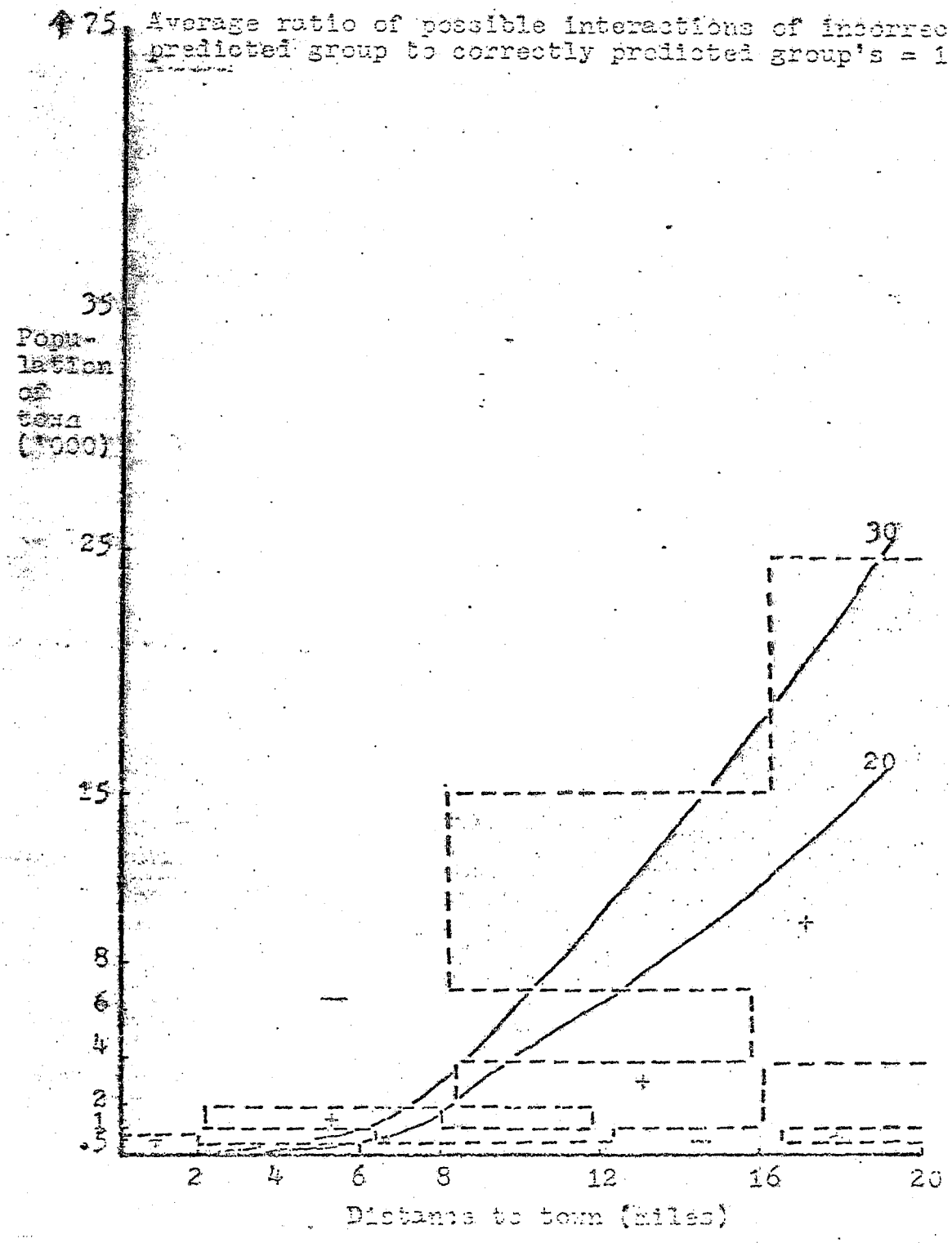


FIGURE 4.4

Zones with the Proportion of Possible Interactions of the Incorrectly Predicted Group Greater or Less Than Average (Grey County, Ontario)



confirms the hypothesis relating model weakness to complex spatial choice situations.

4.3 The Test Repeated with a Different Commodity

A logical extension of the above analyses is to study the performance of the model in explaining spatial allocations of expenditure on a higher order commodity than groceries. Whereas all towns, no matter how small, offer food items for sale, fewer have women's clothing retail outlets. However, it is postulated that, since a consumer's travel-willingness increases for goods available at fewer centers, the probability of complex spatial choice situations is no less than for a more ubiquitously available good. If such is the case the same weakness of the model should appear in a similar analysis to the above.

Employing the indifference surface generated from the households' spatial behavior in relation to maximum women's clothing purchase towns (figure 4.5) a 50% explanation of individual household spatial behavior was obtained. The explained and unexplained groups were compared after the manner employed in 4.1 and 4.2. The relevant graphs (figures 4.6 and 4.7) show much the same relationships between a spatial situation's proportion of incorrect predictions by the model and the related value of the town attractiveness index.

FIGURE A.5

Indifference Surface for Choice of Maximum Women's Clothing Purchase Town by All Households

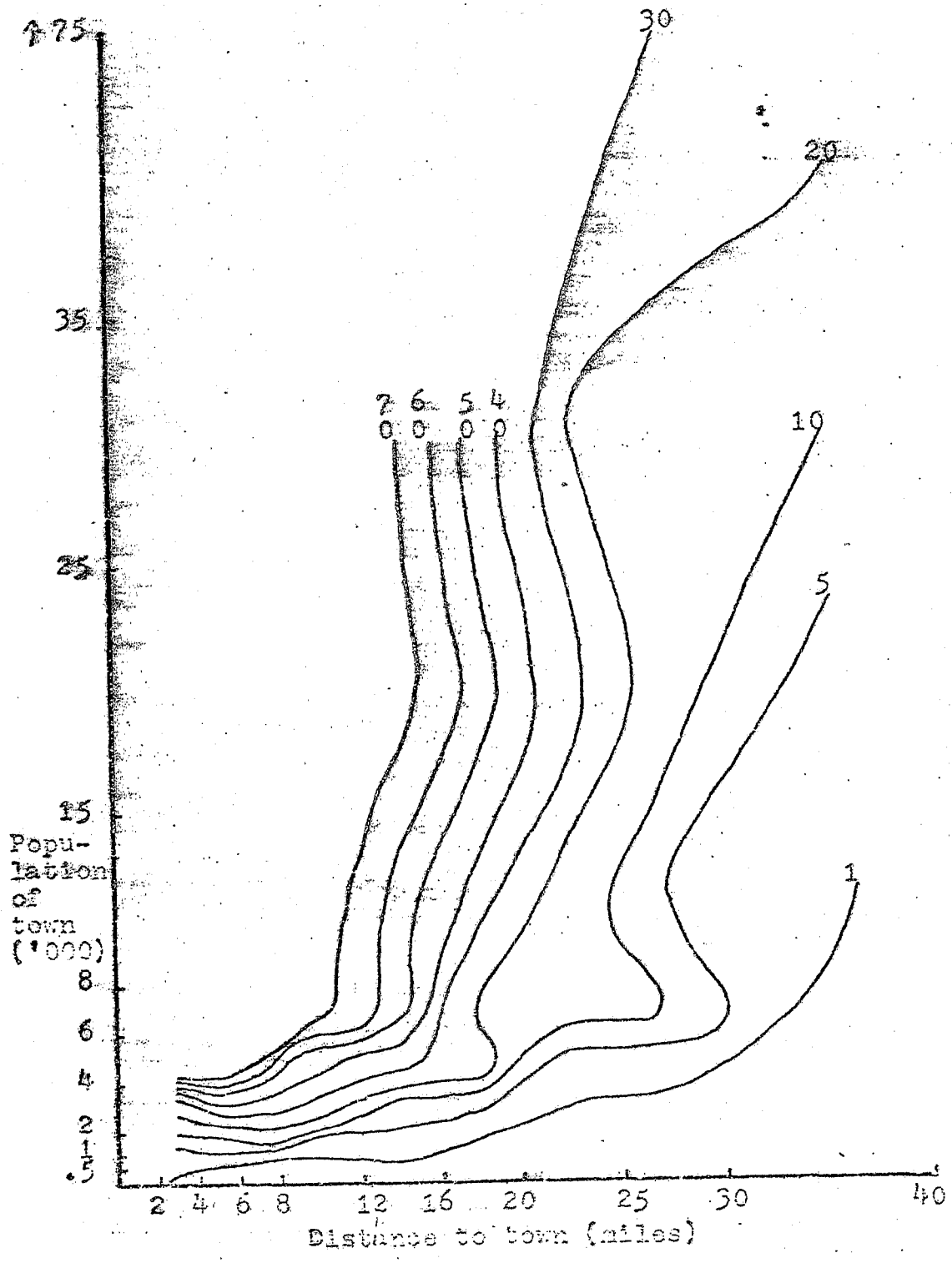


FIGURE 4.6

Zones with the Proportion of Incorrect Predictions Greater or Less Than Average (Women's Clothing Purchases - Iowa)

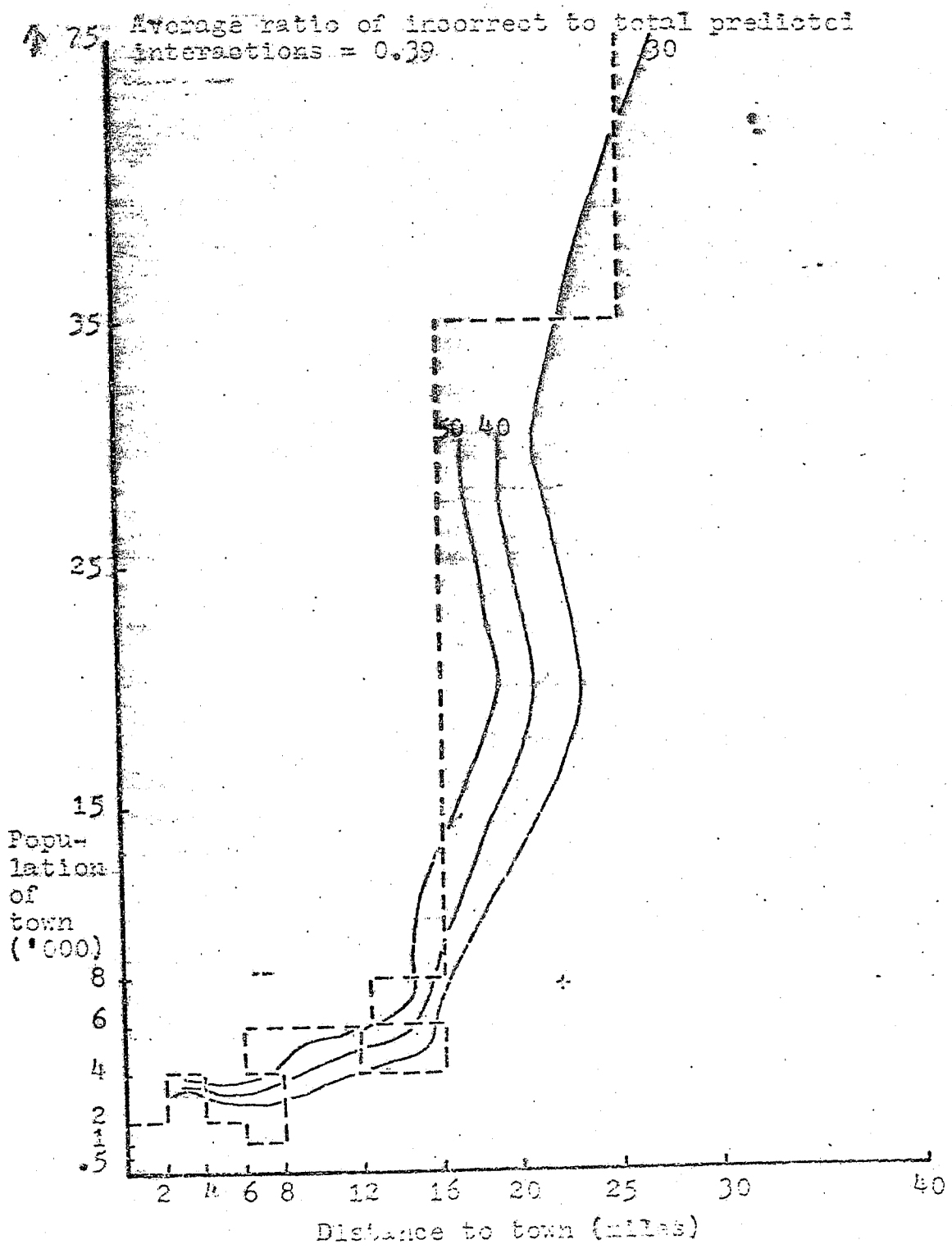
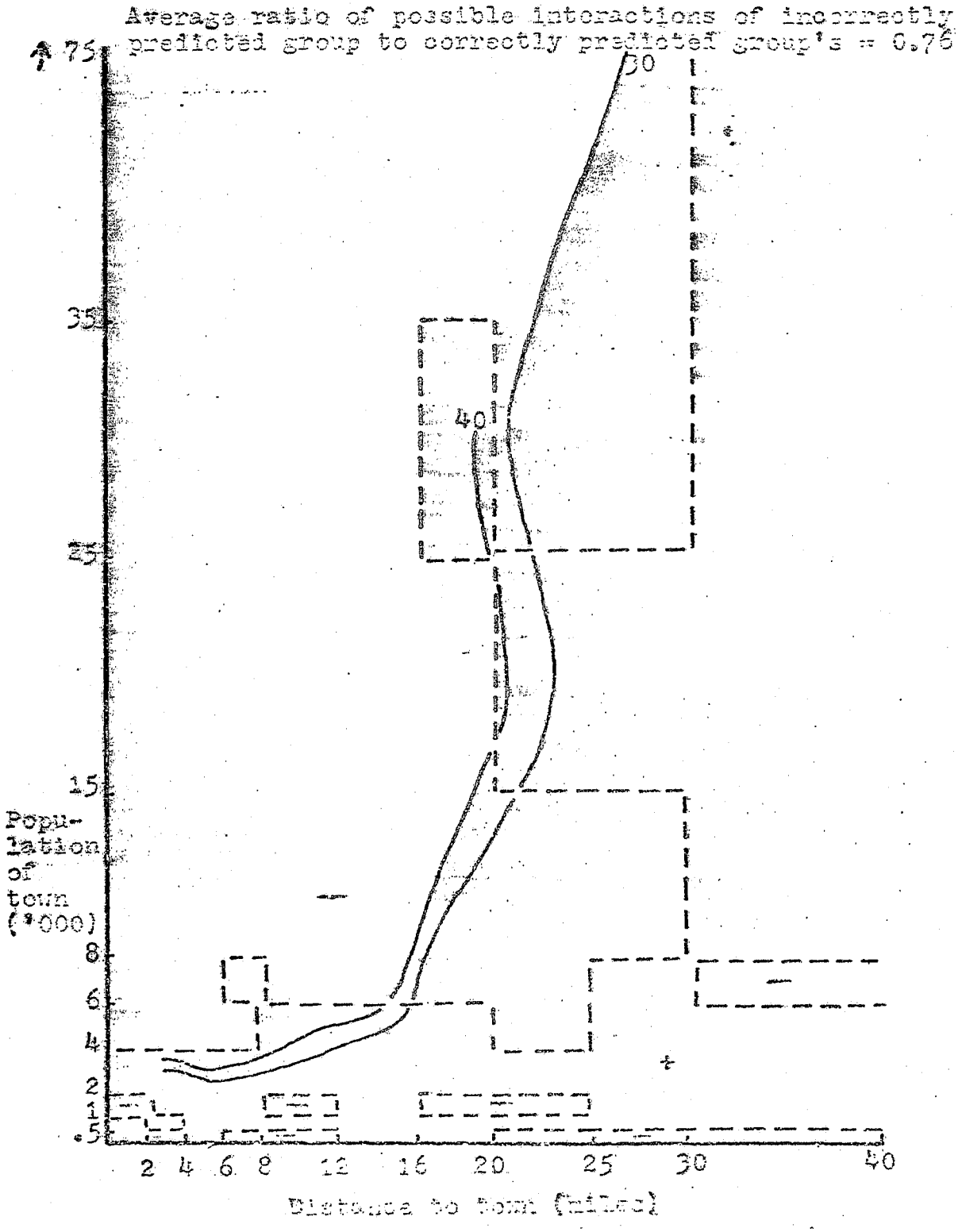


FIGURE 4.7

Zones with the Proportion of Possible Interactions of the Incorrectly Predicted Group Greater or Less Than Average



4.4 The Hypothesis Viewed as Confirmed

The weight of evidence in the above three comparable analyses indicates a similar pattern in the model's inability to explain spatial choice behavior in certain kinds of spatial situations. Where several spatial situations, each with a relatively low proportion of actual to possible interactions, are open to a household, the probability of the model selecting the actual one patronized, is less than in a situation where it is known that a large proportion of possible opportunities was accepted. Three factors which would appear to contribute most to this situation are the inability of a computed indifference surface to replicate an actual surface, the existence of differing space preferences and the deterministic nature of the model. It is in situations of complex spatial choice that the effect of the three error factors is greatest.

4.5 Conclusion

It has been shown that households in complex choice situations are more likely to be wrongly predicted than others and that space preferences and the other two error factors account for these wrong predictions. However, it is very probable that there are cases of households correctly explained which also have dissimilar space preferences, but are capable of being equally well predicted in relatively simple spatial choice situations (i.e. where the most attractive town's index far exceeds

the next most attractive) by any one of several dissimilar indifference surfaces. Thus there is reason to believe that the space preferences of households differ not only between the correctly and incorrectly predicted groups but also within the groups, and that the inaccuracy of the model's indifference surface is only revealed by the more complex choice situations. For this reason and the fact that error can be attributed to factors other than space preferences a replacement is required in future research for the method of analysis which uses the socioeconomic composition of the incorrectly explained group as an indicator of variables possibly related to space preferences. Presently, the only apparent alternative is to categorize the population in terms of high and low "scores" on each variable in turn, and compare the predictive results obtained from each pair of indifference surfaces with the original 65% correct prediction. The contribution to error by factors other than space preferences would be a constant error term in all computations, and thus any significantly increased prediction could be attributed to subdivision of the population on the basis of the variable in question.

CONCLUSION

The primary concern of this study has been to test the hypothesis that differences in the personal characteristics of consumers help explain differences in the spatial choices of urban places for retail expenditure made by consumers similarly located with respect to these centers. A critical examination of previous studies in consumer spatial choice behavior revealed a common weakness. Each assumes that the evaluation by a consumer of alternative spatial opportunities is solely a function of the relative attractiveness or drawing power of urban places, as measured by certain objective indices. Thus the studies reviewed assume a uniform evaluation (ranking) of all conceivable spatial situations by consumers. In all cases, the predictive power of the models in these studies is limited.

It was hypothesized in chapter three, that <u>different</u> consumer rankings of the same spatial opportunities (different space preferences) account for error in existing models of consumer spatial choice behavior. The space preference hypothesis was tested by analyzing the correct and incorrect spatial choice predictions of an indifference curve model, which assumes no differences in consumer space preferences. Those households whose

spatial choices were incorrectly predicted, were hypothesized to have different space preferences from the correctly predicted group. An indifference surface of the "incorrect" households was able to predict 30% of their choice behavior. On this basis, it was inferred that this latter surface represents a set of space preferences different from the original indifference curve model. It was further hypothesized that differences in space preferences are related to differences in the socio-economic characteristics of consumers. This hypothesis was rejected for all fifteen variables considered. Thus the essential conclusion is that, although the existence of space preferences seems to be indicated, they have not been meaningfully related to any measures of personal characteristics.

In chapter four, it was found that the model is in error more often where choices have to be made between relatively similar alternatives. The concentration of error in such situations was explained by the fact that wherever there is a maximum purchase town in combination with a closely ranked alternative, a slight discrepancy between the consumer's expected and actual ranking of spatial alternatives is sufficient to produce an incorrect prediction. It was reasoned that in those spatial situations where spatial opportunities were many, the probability of a maximum purchase town having a closely ranked

alternative was higher than in the spatial situations where spatial opportunities were few. Discrepancies between expected and actual rankings were attributed to the deterministic nature of the model, differing space preferences, and the weakness of the present technique for obtaining an expected indifference surface. Only the contribution of the latter factor to model error has been evaluated with experimental data. The relative contribution to error of each of the other two factors awaits further research. Finally, it was concluded that the technique of comparing the socioeconomic composition of the unexplained group to that of the explained group in order to determine which were the socioeconomic variables most likely related to space preferences was inadequate in view of three factors - the inadequacy of the deterministic model in complex spatial choice situations; its inability to exactly replicate the preference structure of a population characterized by the same space preference structure; and the probability that households with different space preferences are to be found in the correctly predicted group. All three factors tend to negate the assumption used earlier that the sole distinguishing feature between the correctly and incorrectly explained groups is their differing space preferences. Although the existence of space preferences was established, the contrast between the correctly and

incorrectly predicted groups, both in terms of space preferences and their hypothesized socioeconomic correlates, is somewhat weakened by the 'noise' caused by these other error factors.

One direction of future research has been indicated. In addition, a technique for decomposing an aggregate indifference surface into its component surfaces would permit closer study of the differing space preferences represented by these surfaces and provide an alternative approach in the search for consumer characteristics hypothesized to be related to space preferences. The establishment of any such relationship(s) would in turn enable more lawful statements to be made about consumer spatial choice behavior. It would also seem worthwhile to test the effect of using the indifference surface as a probabilistic rather than a deterministic framework for behavioral prediction.

Finally, just as the methodology of the economist's indifference curve has proven a useful tool, there is reason to think that psychological scaling theory and techniques can provide useful theoretical approaches and analytical tools respectively for the study of a dispersed population's scaling of urban places.

APPENDIX 1

Method of Deriving an Indifference Surface from Spatial Choices and Possible Opportunities of All Households

The model postulates that in a household's choice of maximum grocery purchase town from a set of alternative possible towns, the two most significant variables are, town population and distance from household to town. The former variable, as has been empirically proven[1], is strongly correlated with a town's retail functional diversity and its breadth of merchandise, whilst the friction of distance has been frequently demonstrated to be a powerful agent in the spatial dimension of economic systems. The consumer's choice is reasoned as being a function of his evaluation of the different alternative spatial interactions open to him. The model, by definition, implies that this consumer evaluation has a regular pattern

[1] Stafford, H., "The Functional Bases of Small Towns", Economic Geography, Volume 39, Number 2, 1963, pp. 165 - 175.
Thomas, E. N., "Some Comments on the Functional Bases for Small Iowa Towns", Iowa Business Digest, Volume 31, Number 2, Bureau of Business and Economic Research, University of Iowa, Iowa City, 1960.
Berry, B. J. L., Barnum, H. G. and Tennant, R. J., "Retail Location and Consumer Behavior", Papers and Proceedings of the Regional Science Association, Volume 9, 1962, pp. 65 - 106.

which can be lawfully described in terms of the two above mentioned variables.

The necessary data to test the model and its reasoning are as follows:

1. The locational co-ordinates of the 456 rural households in the sample survey.
2. The locational co-ordinates of all towns of over 55 persons in the state of Iowa, together with their populations.
3. The actual town which each household chose as its maximum grocery purchase town.

"A consumer's spatial situation is defined as all distance and town-size combinations which surround him"[2], the calculation of which is facilitated by the data under headings 1 and 2. Each consumer's spatial situation, therefore, is made up of a set of alternative central places. The model attempts to predict for each household, the highest ranked town in terms of grocery dollar expenditure.

The method used by the model entails knowing, for each arbitrarily defined town-size and distance combination, the sum of all actual interactions with each combination, as well as the sum of all possible interactions with each combination (see diagram A 1.1).

[2] Rushton, G, op. cit., p. 28.

DIAGRAM A 1.1

Actual and Possible Interactions
of Every Household

1 Household	2 Actual Maximum Grocery Purchase Town	3	4 Possible Household/Town Interactions	5
	Population	Distance	Population	Distance
1	P_3	$D_{1,3}$	P_1 P_2 . . . P_j	$D_{1,1}$ $D_{1,2}$ . . . $D_{1,j}$
2	P_{58}	$D_{2,58}$	P_1 P_2 . . . P_j	$D_{2,1}$ $D_{2,2}$ . . . $D_{2,j}$
. . . n	P_{92}	$D_{n,92}$	P_1 P_2 . . . P_j	$D_{n,1}$ $D_{n,2}$ . . . $D_{n,j}$

where P_j = population of jth town

$D_{n,j}$ = distance from n^{th} household to j^{th} town

If P_1D_k = 1th town population category/k^{th} distance category combination,

then for each of the 1 x k town population/distance combinations, the appropriate values, within the population and distance limits of that combination, in columns 2, 3, 4 and 5 are summed. Hence a matrix of actual and possible interactions is obtained (see diagram A 1.2).

DIAGRAM A 1.2

Matrix of Actual and Possible Interactions

	$AI_{1,1}$ $PI_{1,1}$	$AI_{1,2}$ $PI_{1,2}$					$AI_{1,k}$ $PI_{1,k}$
	$AI_{2,1}$ $PI_{2,1}$						
Town Population							
	$AI_{1,1}$ $PI_{1,1}$						$AI_{1,k}$ $PI_{1,k}$

Distance

where $AI_{1,k}$ = number of actual interactions in 1 th town population and k th distance category,

$PI_{1,k}$ = number of possible interactions in 1th town population and k^{th} distance category, and

$AI_{1,k}/PI_{1,k}$ = ordinal value describing hypothesized attractiveness of that spatial situation in relation to others.

It is reasoned that where a large number of possible interaction opportunities are accepted by households, i.e. where the AI/PI ratio is high, that towns in such spatial situations are more attractive than those towns in situations where few of the possible interaction opportunities are accepted. Deduced from this initial proposition is the hypothesis that a household will be indifferent between two or more towns in spatial situations where the AI/PI ratio (known as the "town attractiveness index") has the same value. On this basis a set of indifference curves were drawn (see diagram A 1.3) over the AI/PI matrix, joining points of equal value.

DIAGRAM A 1.3

Indifference Surface Fitted To Actual/Possible Interactions Matrix

100	90	80	60	50	40	30 20	0
90	80	74	59	45	35	25	5
84	76	65	55	43	32	20	12
80	70	60	50	40	25	15	10
72	65	55	45	35	20	15	5
70	60	48	32	22	16	5	1
60	42	32	22	14	6	1	0
30	24	17	10	7	3	0	0

Indifference curve value = $\frac{AI}{PI} \times 100$

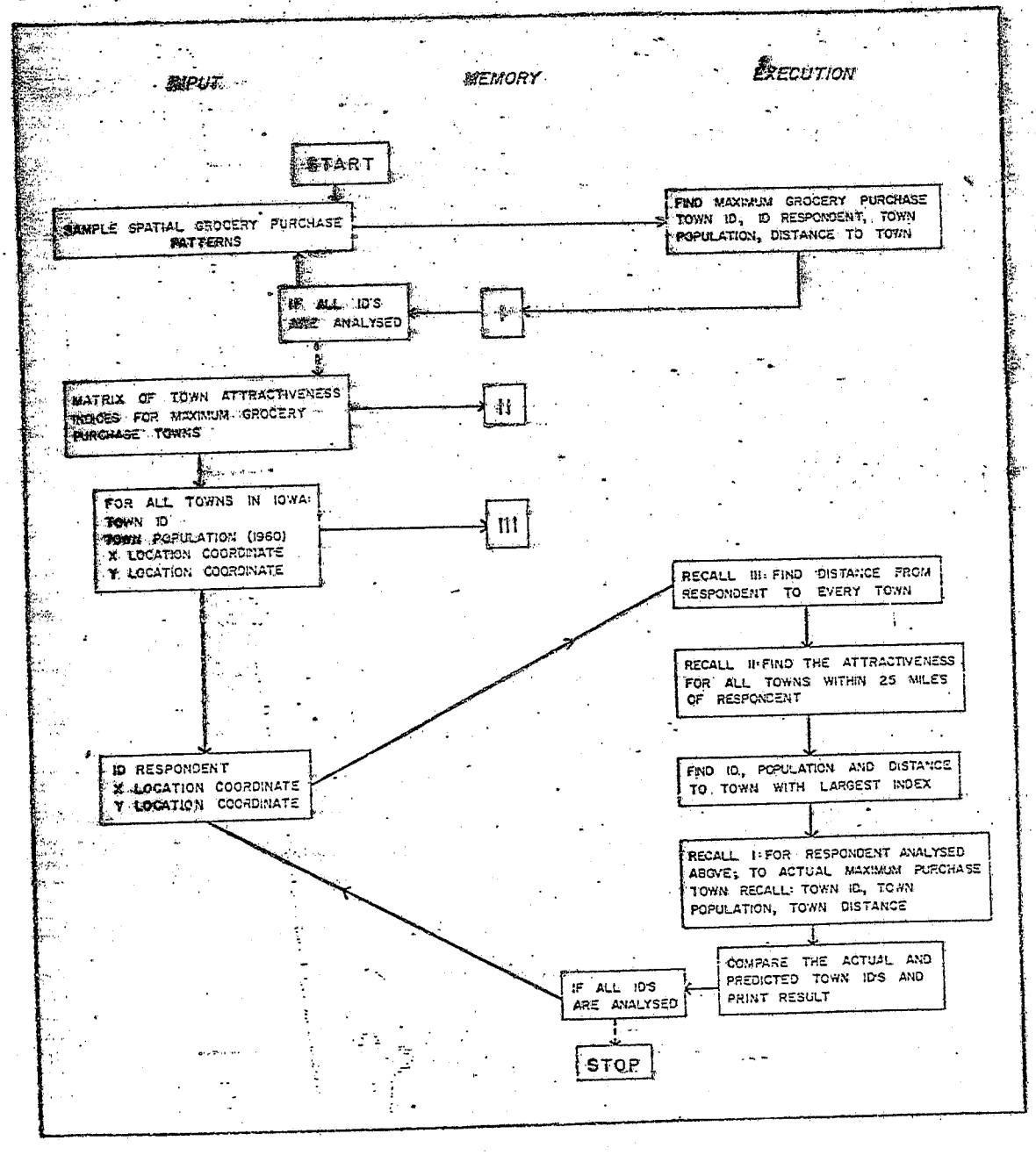
Values in each cell is arbitrarily regarded as value for mid-point of cell.

On the basis of this surface, it is argued that a household will choose as its maximum purchase town, that alternative available to it, lying at the highest point on the indifference surface, and that it will be indifferent between two or more towns satisfying this condition. In fact, the original ratios, which stated the frequency of interactions in each spatial situation, can be regarded as statements of the relative attractiveness of a possible interaction. Being a deterministic model, however, the most "attractive" town, i.e. the predicted town, automatically assumes the probability 1.0 of being the maximum purchase town. The prediction for each household is then compared with actual behavior and the households classed by correct (66%) and incorrect (34%) prediction. Diagram A 1.4[3] illustrates the entire program sequence for testing the model.

[3] Rushton, G., op. cit., p. 40.

DIAGRAM A 1.4

Test of the Model - The Program Sequence



Source: Rushton, G., op. cit., p. 40

APPENDIX 2

A Test of the Efficiency of the Indifference Surface in Replicating Known Space Preference Structures

Rushton performed a series of tests of the hypothesis that a ranking of spatial situations by the values of town attractiveness indices for an indifference surface will be identical to the ordering of the same spatial situations by decision makers. It is this ordering which leads to their observed behavior. In the test, using experimental data, all consumers were assigned the same consistent behavior patterns, the rules of which were predetermined. Therefore, knowing the spatial choice each consumer would make in accordance with the given rules, an indifference surface was generated from this behavior. To test that this surface was identical to the space preference structure of the population, the surface was used to predict each consumers behavior and this expected behavior was then compared with the observed hypothetical behavior. A one hundred percent accurate prediction would indicate that the surface did, in fact, correspond to the space preference structure of the group for the same spatial situations. The results of the test, which was repeated using three different "rules of spatial behavior" are shown in table A 2.1.

TABLE A 2.1

Test of the Efficiency of the Computed Indifference Surface

	Hypothesized Spatial Behavior Patterns		
	Nearest Town Patronized	Town Patronized, Having Largest Value of $\frac{T_j}{d_{ij}}$ *	Town Patronized, Having Largest Value of $\frac{T_j}{d_{ij}^2}$ *
Percent of 480 sample households sent to same town by both rankings of spatial situations	91.2	90.0	84.6

Source: Rushton, G., "Analysis of Spatial Behavior by Revealed Space Preference", Research Report, Computer Institute for Social Science Research, Michigan State University, October, 1967.

* where T_j = population of jth town,

and d_{ij} = distance between household i, and town j.

APPENDIX 3

Tables of Town Attractiveness Indices from which Indifference Surfaces are Derived and Tables of Possible Interactions

TABLE A 3.1

a) Town Attractiveness Indices for Total Sample Population

Town Population	Miles from Household to Town							
	0	2	4	6	8	12	16	20
500	37.5	12.3	7.2	2.9	0.3	0.0	0.1	
1,000	12.5	47.1	24.3	19.1	3.4	0.6	0.2	
2,000	60.0	80.0	60.0	29.8	17.5	1.6	0.0	
4,000	100.0	75.0	88.2	45.8	38.6	11.5	0.8	
6,000	-	100.0	100.0	73.3	45.8	25.0	9.5	
8,000	-	100.0	100.0	100.0	69.2	37.5	15.2	
15,000	-	100.0	100.0	87.5	33.3	50.0	11.8	
25,000	-	100.0	100.0	-	45.5	87.5	7.7	
35,000	-	-	100.0	100.0	50.0	43.8	28.6	
75,000	-	0.0	25.0	100.0	100.0	50.0	0.0	

Note:- Dash denotes no possible interactions were recorded for given spatial situation.

b) Town Attractiveness Indices for Unexplained Group

Town Population	Miles from Household to Town							
	0	2	4	6	8	12	16	20
500	55.6	23.4	18.3	8.0	0.8	0.0	0.2	
1,000	0.0	57.1	25.8	25.0	7.1	1.7	0.6	
2,000	0.0	0.0	23.1	21.7	16.0	4.4	0.0	
4,000	-	0.0	0.0	0.0	21.9	14.3	2.9	
6,000	-	-	-	0.0	0.0	17.4	11.1	
8,000	-	-	-	-	20.0	10.0	22.2	
15,000	-	-	-	0.0	20.0	37.5	11.1	
25,000	-	-	-	-	0.0	50.0	0.0	
35,000	-	-	-	-	0.0	37.5	50.0	
75,000	-	-	0.0	-	100.0	0.0	0.0	

TABLE A 3.2

a) Town Attractiveness Indices for Households with 0 to 13 Years Residence in Present House

Town Population	Miles from Household to Town							
	0	2	4	6	8	12	16	20
500	33.3	11.5	6.1	2.7	0.2	0.0	0.3	
1,000	0.0	31.3	19.4	17.4	2.4	0.5	0.0	
2,000	50.0	75.0	56.3	28.1	15.0	1.0	0.0	
4,000	100.0	75.0	100.0	26.7	34.1	15.0	0.0	
6,000	-	100.0	100.0	71.4	50.0	26.3	18.8	
8,000	-	100.0	100.0	100.0	78.6	55.0	12.5	
15,000	-	100.0	100.0	85.7	66.7	62.5	10.0	
25,000	-	-	100.0	-	60.0	100.0	20.0	
35,000	-	-	100.0	100.0	100.0	41.7	42.9	
75,000	-	0.0	50.0	100.0	-	50.0	0.0	

b) Town Attractiveness Indices for Households with 14 to 75 Years Residence in Present House

Town Population	Miles from Household to Town							
0	0	2	4	6	8	12	16	20
500	40.0	12.5	8.2	2.9	0.8	0.2	0.1	
1,000	50.0	61.1	28.6	20.5	4.9	0.6	0.5	
2,000	66.7	87.5	65.0	31.3	20.3	2.2	0.0	
4,000	100.0	75.0	81.8	77.8	44.8	8.0	2.0	
6,000	-	100.0	100.0	75.0	50.0	28.6	3.8	
8,000	-	100.0	-	100.0	58.3	15.4	16.7	
15,000	-	-	-	100.0	14.3	33.3	14.3	
25,000	-	100.0	100.0	-	33.3	80.0	20.0	
35,000	-	-	-	100.0	42.9	33.3	42.9	
75,000	-	-		-	100.0	50.0	0.0	

TABLE A 3.3

a) Possible Interactions of Explained Group

Town Population	Miles from Household to Town							
	0	2	4	6	8	12	16	20
500	15	67	126	179	524	761	933	
1,000	5	20	39	49	151	229	285	
2,000	3	16	22	24	76	117	165	
4,000	2	7	16	17	38	52	88	
6,000	0	4	5	12	17	17	24	
8,000	0	2	5	10	21	22	15	
15,000	0	5	2	7	4	6	8	
25,000	0	1	5	0	6	6	8	
35,000	0	0	2	2	6	8	8	
75,000	0	1	3	2	0	2	2	

b) Possible Interactions of Unexplained Group

Town Population	Miles from Household to Town							
0	0	2	4	6	8	12	16	20
500	9	47	82	100	260	403	511	
1,000	3	14	31	40	112	118	162	
2,000	2	4	13	23	50	68	78	
4,000	0	1	1	7	32	35	34	
6,000	0	0	0	3	7	23	18	
8,000	0	0	0	0	5	10	18	
15,000	0	0	0	1	5	8	9	
25,000	0	0	0	0	5	2	5	
35,000	0	0	0	0	2	8	6	
75,000	0	0	1	0	1	2	2	

APPENDIX 4

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

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

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

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

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

Bibliography

Berry, B. J. L., Geography of Market Centers and Retail Distribution, Prentice-Hall, Englewood Cliffs, N. J., 1967.

Berry, B. J. L., Barnum, H. G., and Tennant, R. J., "Retail Location and Consumer Behavior", Papers and Proceedings of the Regional Science Association, Volume 9, 1962, pp. 65 - 106.

Christaller, W., Central Places in Southern Germany, translated by Carlisle W. Baskin, Prentice-Hall, Englewood Cliffs, N. J., 1966.

Edwards, W., "The Theory of Decision Making", Psychological Bulletin, Volume 51, Number 4, 1954, pp. 380 - 417.

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

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

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

Huff, D. L., "A Topographic Model of Consumer Space Preferences", Papers and Proceedings of the Regional Science Association, Volume 6, 1960, pp. 159 - 173.

Huff, D. L., "A Probabilistic Analysis of Consumer Spatial Behavior", Emerging Concepts in Marketing, Proceedings of the Winter Conference of the American Marketing Association, W. S. Decker (ed.), December, 1962.

Isard, W., Location and Space-Economy, Regional Science Studies Series, 1, Cambridge, Mass., 1962.

Bibliography (Continued)

Johassen, C. T., The Shopping Center versus Downtown, Bureau of Business Research, Ohio State University, Columbus, 1955.

Malm, R., Olsson, G. and Wärneryd, O., "Approaches to Simulations of Urban Growth", Geografiska Annaler, Volume 48B, Number 1, 1966, pp. 9-22.

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

Mertz, W. L., and Hamner, L. B., "A Study of Factors Related to Urban Travel", Public Roads, Volume 29, April, 1957, pp. 208 - 212.

Michelson, W., "An Empirical Analysis of Urban Environmental Preferences", Journal of the American Institute of Planners, volume 32, number 6, 1966, pp. 355 - 360.

Miller, G. A., "The Study of Intelligent Behavior", Annals of the Computation Laboratory of Harvard University, volume 31, 1962, pp. 6 - 15.

Mitchell, R. A., "An Explanation of the Expenditure Pattern of a Dispersed Population", unpublished Ph. D. dissertation, State University of Iowa, 1964.

Murdie, R. A., "Cultural Differences in Consumer Travel", Economic Geography, volume 41, number 3, July, 1965, pp. 211 - 233.

Pareto, V., Manuale di economia politica, con una introduzione ulla sciensa sociale, Milan, 1906.

Reilly, W. J., The Law of Retail Gravitation, The Knickerbocker Press, New York, 1931.

Rushton, G., Spatial Pattern of Grocery Purchases by the Iowa Rural Population, Studies in Business and Economics, New Series Number 9, The University of Iowa, Iowa City, 1966.

Bibliography (Continued)

Rushton, G., "Analysis of Spatial Behavior by Revealed Space Preference", Research Report, Computer Institute for Social Science Research, Michigan State University, East Lansing, October, 1967.

Rushton, G., Golledge, R. G., and Clark, W. A. V., "Formulation and Test of a Normative Model for the Spatial Allocation of Grocery Expenditures by a Dispersed Population", Annals of the Association of American Geographers, volume 57, number 2, June, 1967, pp. 389 - 400.

Siegel, S., Siegel, A. E. and Andrews, J., Choice, Strategy and Utility, McGraw-Hill, New York, 1964.

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

von Neumann, J. and Morgenstern, O., Theory of Games and Economic Behavior, Princeton University Press, Princeton, 1944.