HOUSEHOLD SOCIO-ECONOMIC AND DEMOGRAPHIC CHARACTERISTICS AS DETERMINANTS OF FOOD EXPENDITURE BEHAVIOR

Thesis for the Degree of Ph.D.
MICHIGAN STATE UNIVERSITY

Robert Omer Herrmann

1964

THESIS

This is to certify that the

thesis entitled

HOUSEHOLD SOCIO-ECONOMIC AND DEMOGRAPHIC CHARACTERISTICS AS DETERMINANTS OF FOOD EXPENDITURE BEHAVIOR

presented by

Robert Omer Herrmann

has been accepted towards fulfillment of the requirements for

Ph.D. degree in Agricultural Economics

James T Bonnen

Major professor

Date January 7, 1964

O-169

LIBRARY
Michigan State University

82 279
Rec'd 2 1 1994
M 139

ABSTRACT

HOUSEHOLD SOCIO-ECONOMIC AND DEMOGRAPHIC CHARACTERISTICS AS DETERMINANTS OF FOOD EXPENDITURE BEHAVIOR

by Robert Omer Herrmann

Despite the wide current interest in the effects of population composition on economic growth and development, relatively little empirical work has been done on the relationship between population composition and consumer demand. Before the aggregate effects of changes in population composition can be determined, estimates are needed of the effect of particular household demographic and socio-economic characteristics on the demand for individual commodities. In order to derive such estimates an economic model including socio-economic and demographic variables, as well as economic variables, is required. This study focusses on the problems of developing such a model to explain household food expenditures.

The first problem dealt with is that of specifying a structural model explaining household food expenditure. Such a model identifies the dependent variable and the independent variables which affect it, but not the form of their relationship. The eight socio-economic and demographic variables examined to determine whether they should be included in the structural model were urbanization, region, race, stage in the family life cycle, social class, education of the homemaker, self-employment status of the household head and employment status of the homemaker. Individual household data selected from the sample of the U. S. Depattment of Agriculture 1955 Household Food Consumption Survey were employed in the analysis. Eight separate regressions were estimated.

In each, the various categories of the particular characteristics were represented by dummy variables. The statistical significance of each of the characteristics was evaluated with an F test. A test of economic significance also was applied.

Because of the magnitude of the effects of urbanization and region, certain additional tests were made in which one or both of these variables were controlled. After these tests, it was concluded that the structural model should include variables to represent the effects of urbanization, region, stage in the family life cycle, social class and education of the homemaker.

The second major group of problems in working toward an economic model of household food expenditure were ones of variable specification. An examination of the estimates of the effects of the characteristics suggested that representation by continuous variables did not appear warranted. None of the variables displayed a clear pattern of effect which could be represented by a single continuous variable. For this reason, the use of dummy variables appears preferable.

Third problem dealt with was that of model specification. A classification of alternative models of the effects of household characteristics on food expenditure behavior was developed:

1) models which assume that the effects of a given characteristic can be represented by one or more independent variables,

2) models which assume that a characteristic affects food expenditure only through its effect on the income-expenditure relationship,

3) models which assume that a given characteristic affects food expenditure through its effect on the relationships of the independent variables to food expenditure.

It was concluded that a single model incorporating the effects of a number

of socio-demographic characteristics is feasible. The general form of such a model is suggested.

The estimates of the effects of urbanization and region were employed to calculate the effects of changes in the distribution of households among urbanization and region categories in the 1950-1960 decade. It was concluded that changes by urbanization produced an important increase in average food expenditure. Regional shifts, produced a somewhat smaller increase. The sizable effects of these shifts in population composition indicate that the completion of a single economic model including all the significant household socio-demographic characteristics and the estimation of the parameters of such a model would be desirable and useful.

HOUSEHOLD SOCIO-ECONOMIC AND DEMOGRAPHIC CHARACTERISTICS AS DETERMINANTS OF FOOD EXPENDITURE BEHAVIOR

By

Robert Omer Herrmann

A THESIS

Submitted to
Michigan State University
In partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics

1964

g 3 15 88
2-5-65

ACKNOWLEDGMENTS

The author wishes to express his thanks to the many people who advised and assisted him in the course of this study.

Dr. James T. Bonnen was a source of encouragement and advice throughout the study. His standards of academic excellence were and will continue to be a source of inspiration to the author.

Dr. Lester V. Manderscheid counselled the author on a number of statistical problems which arose in the course of this study. His advice helped the author traverse the sometimes difficult path between statistical theory and empirical analysis.

Dr. Robert L. Gustafson and Dr. James D. Shaffer also provided useful advice and suggestions.

The author wishes to acknowledge the help of Richard Clark in coding the occupational data. The departmental Statistical Pool provided cheerful assistance throughout the study. The author wishes to acknowledge, in particular, the assistance of Mrs. Laura Flanders, who handled the computer production for this study.

The author wishes to thank Dr. L. L. Boger and the Department of Agricultural Economics for their financial assistance, for the use of departmental facilities and for their friendly encouragement throughout the course of his graduate studies.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	2
	Estimates of the Effects of Socio-economic and Demographic Characteristics on Expenditure Behavior.	2
	Procedure.	6
II.	SPECIFYING THE STRUCTURAL MODEL	9
	Selection of Socio-demographic Variables for Inclusion in the Structural Model	9
	Selection of a Preliminary Economic Model.	20
	The Tests of Significance Employed	30
	Analysis and Results	34
	Determination of the Effects of Correlated Biases.	54
	Conclusions about the Specification of the Structural Model.	62
III.	SPECIFICATION OF SOCIO-ECONOMIC AND DEMOGRAPHIC VARIABLES	66
	Examination of the Effects of the Variables.	66
	Implications of the Observed Patterns of Effect for the Quantification of the S-D Variables.	100
IV.	ALTERNATIVE ECONOMIC MODELS OF THE EFFECTS OF S-D VARIABLES	102
	Levels of Measurement and the Models	103
	Classification of the Alternative Models	107
	Some Conclusions about Economic Model Specification.	117
V.	CONCLUSIONS AND IMPLICATIONS	122
	Conclusions.	123
	Implications	126
VI.	APPENDIX	135
VII.	BIBLIOGRAPHY	159

LIST OF TABLES

TABLE		Page
2.1	Characteristics Explaining Differences in Food Expenditure Behavior between Households	12
2.2	Alternative Models Examined in Selecting a Preliminary Economic Model	25
2.3	Tests of Statistical Significance	37
2.4	Tests for the Effects of Correlated Biases	57
2.5	Estimated Adjusted Coefficients of Multiple Determination ($\bar{R}^2$)	64
3.1	Regression Results for Urbanization and Region	70
3.2	Effects of Intercorrelations with Region on the Coefficients Estimating Urbanization Effects	72
3.3	Effects of Intercorrelations with Urbanization on the Coefficients Estimating Regional Effects	74
3.4	Regression Results for Race	75
3.5	Effects of Intercorrelations with Urbanization and Region on the Coefficient Estimating Race Effects	77
3.6	Descriptions of the Life Cycle Categories with Number of Observations in Each	78
3.7	Regression Results for Stage in the Family Life Cycle	79
3.8	Descriptions of the Social Class Categories Employed with Number of Observations in Each	83
3.9	Regression Results for Social Class	84
3.10	Descriptions of the Categories of Education of the Homemaker Employed with Number of Observations in Each	87
3.11	Regression Results for Education of the Homemaker	89
3.12	Regression Results for Self-Employment of the Household Head	92
3.13	Regression Results for Employment of the Homemaker	93

3.14	Regression Results for Interaction Categories of Region and Urbanization	95
3.15	Tests of the Significance of Differences between Regions at Given Levels of Urbanization	97
5.1	Changes in the Distribution of Households by Urbanization and Resultant Changes in Expenditure, 1950 to 1960	128
5.2	Changes in the Distribution of Households by Region and Resultant Changes in Expenditure, 1950 to 1960	130
5.3	Changes in the Distribution of Households by Age of Head, 1950 to 1960	132
II .1	Cost of One Week's Food Under the Moderate-Cost Food Plan and Consumption Equivalence Scale Coefficients for Thirteen Age-Sex Categories	145
III .1	Socio-Economic Index Scores of Typical Occupations Included in the Four Social Class Categories Employed	156
III .2	Distribution of Households between Social Class Categories in This and Previous Studies	157

LIST OF FIGURES

FIGURE		Page
2.1	The Multi-Track System of Family Life Cycle Stages Employed	43
3.1	Stages in the Family Life Cycle and Estimates of Their Effects on Food Expenditure	80

LIST OF APPENDICES

APPENDIX		Page
I	The Data Employed	135
II	The Standard Meal Units Variable	143
III	The Social Class Variable	146

CHAPTER I

Introduction

In recent years an increasing amount of attention has been given to changes in the composition of the population and the impact of these changes on economic growth and development. Despite the wide interest in this subject, there have been relatively few studies which have attempted to estimate the effect of changes in population composition on consumer demand.[1] One reason for the slow development of such studies appears to be the lack of suitable estimates of the effects of household socio-economic and demographic characteristics on expenditure behavior. Such estimates are, of course, an essential prerequisite to any evaluation of the effects of changes in the distribution of population characteristics on aggregate consumer expenditures.

The estimation of the effects of household socio-economic and demographic variables implies the need for an economic model of food expenditure behavior, which incorporates both economic variables and variables representing relevant household socio-economic and demographic (S-D) characteristics. This study will focus on the problems involved in the development of such an economic model.

Estimates of the Effects of Socio-economic and Demographic Characteristics on Expenditure Behavior

A variety of techniques have been used to estimate the effects of household S-D characteristics. Few of the estimates produced have been

---

[1] One of the major groups of papers on this subject is National Bureau of Economic Research, Demographic and Economic Change in Developed Countries ("Special Conference Series", Vol. 11; Princeton: Princeton University Press, 1960). The effects of household characteristics on expenditures are taken into account in many of the papers included in Irwin Friend and Robert Jones (eds.), Proceedings of the Conference on Consumption and Saving (Philadelphia: University of Pennsylvania, 1960).

suitable for estimating the effects of changes in population composition on consumer expenditures, however. The techniques used to produce the presently available estimates of the effects of household S-D characteristics and the reasons that these presently available estimates are of only limited usefulness will be discussed in the following two sections.

Estimating Techniques

The techniques used to estimate the effects of household S-D characteristics have been of two principal types. Regression and analysis of variance techniques have been employed to obtain some of the estimates. Other estimates have been produced through the use of multiple cross-classifications of household expenditure survey data to isolate the effects of a particular characteristic.

Several types of economic models designed to permit estimation of the effects of S-D characteristics appear to be feasible. The most widely used such models have been ones developed for use with cross-sectional micro-level data. Among the studies of this type are Charles Zwick's study of the effects of household characteristics on income and price elasticities for meat and Jean Crockett's study of the effects of household characteristics on expenditures for food for home consumption.[1] The effects of household characteristics also have been estimated in a time-series micro-level model. Such a model was employed by Willard Sparks

---

[1] Charles Zwick, "Demographic Variation: Its Impact on Consumer Behavior," Review of Economics and Statistics, Vol. 39 (November, 1957), pp. 451-56.
Jean Crockett, "The Demand Relationships for Food," Proceedings of the Conference on Consumption and Saving, Vol. I, Irwin Friend and Robert Jones (eds.), (Philadelphia: University of Pennsylvania, 1960), pp. 293-310.

in his analysis of Lansing consumer panel data.[1] One model employed by Sparks includes an individual constant term for each individual household. The estimate of this term is an estimate of persistent differences in the behavior of the family over time. The effect of particular S-D characteristics on the constant term estimates obtained by Sparks requires further study. A third method of estimating the effects of S-D characteristics suggests itself. This is the use of macro-level time-series data on expenditures, incomes and population composition to estimate the effects of changes in population. Such a method would relate observed changes over time in aggregate or per capita expenditure to observed changes in population composition.

Limitations of Past Estimates

Many of the past estimates of the effects of S-D characteristics are of limited usefulness in estimating the effects of changes in population composition either because of the estimation techniques employed or the data utilized.

Several of the past studies which have estimated the effects of household characteristics have been based on single city samples. Others have employed a sample including only urban households. Such studies cannot provide complete information on the key variables, urbanization and region.

The usefulness of some of the estimates of effects obtained by regression and analysis of variances is limited by the failure, in most studies, to control the effects of correlated characteristics while estimating the effect of particular characteristics on expenditure.

---

[1] Willard Robert Sparks, "Estimates of the Demand for Food from Consumer Panel Data," (unpublished Ph.D. dissertation, Department of Agricultural Economics, Michigan State University, 1961), pp. 43-44.

Household characteristics are not distributed independently. As we shall see in Chapter II, failure to control the effects of correlated characteristics may affect the estimates obtained to a marked extent.

The two micro-level techniques of estimating the effects of household S-D characteristics appear to be capable of producing reliable estimates of these effects. The macro-level technique does not appear likely to produce any usable estimates, however. Multi-collinearity, lack of independence between successive observations and the small number of observations available limit the usefulness of this latter technique.

Multiple cross-classifications of household expenditure survey data also have been employed to isolate the effect of household characteristics on expenditure. Such cross-classifications permit us to determine the differences in expenditure between households falling into different categories of one characteristic with the effects of one or more other characteristics held constant.

Cross-classification has been used in the analysis of the data from three recent major studies of household expenditures. These studies include the Bureau of Labor Statistics 1950 Survey of urban households, the Department of Agriculture 1955 Household Food Consumption Survey and the LIFE Study of Consumer Expenditures.[1] The data in these three studies have been tabulated with up to four-way cross-classifications. The usefulness of these tabulations is limited by the lack of information

---

[1] University of Pennsylvania, Study of Consumer Expenditures, Incomes and Savings, Vol. III.: Expenditures for Food, Beverages and Tobacco. (Philadelphia: University of Pennsylvania, 1956). U. S. Department of Agriculture, Food Consumption of Households in the United States "Household Food Consumption Survey 1955," Report No. 1 (Washington: USGPO, 1956). LIFE, LIFE Study of Consumer Expenditures, Vol. I. (New York: Time, Inc., 1957).

on intra-cell variance. Such information is necessary in order to determine the significance of differences between cell means. The usefulness of these tabulations also is limited by the number of cross-classifications employed. A maximum of four cross-classifications have been used because of the limited number of household observations available. This permits the holding of three factors constant, while the effect of the fourth is studied. After income and family size have been held constant, only one other variable can be held constant in a four-way cross classification. This may not prove to be sufficient in analyses where it would be desirable to hold income, family size, region and urbanization all constant in order to study the effect of a fifth characteristic. Cross classification techniques have the further disadvantage that they waste information by treating variables such as income as categories rather than as continuous variables.

It appears that what is needed for estimating the effects of changes in population composition are estimates of the effects of S-D characteristics free from correlated biases and based on nationwide data. This study will work in the direction of an economic model designed to produce such estimates.

Potential Uses of Estimates

Better estimates of the effects of household socio-economic and demographic (S-D) characteristics on expenditures clearly would be of substantial value in economic forecasting applications. Coupled with projections of the distribution of S-D characteristics, estimates of the effects of these characteristics permit estimation of the impact of changes in population composition on aggregate expenditure. Estimates of the relationship of S-D characteristics to expenditure behavior also are needed for the development of simulation models of household expenditure

behavior. The development of estimates of the separate effects of different household characteristics appears to be one of the principal problems in estimating the probability a given household will engage in a particular type of economic act.[1] Such probabilities are a key variable in simulation models of the type developed by Guy Orcutt and his colleagues.

While estimates of the effects of household S-D characteristics would be useful in economic forecasting applications, the economic model used to produce these estimates would have certain additional uses. A fully developed economic model relating economic, socio-economic and demographic variables to food expenditure, or other categories of expenditure, would serve to integrate economic and sociological variables. Such a model along with estimates of its parameters would provide an empirical foundation for theories of consumer demand incorporating variables from sociology as well as from economics.

Procedure

This study will undertake to resolve certain of the problems involved in the development of an economic model of food expenditure behavior including both economic variables and variables representing household socio-economic and demographic characteristics. The steps taken in this study toward the development of this model and the scope of the study are discussed in the following sections.

---

[1] Guy H. Orcutt, Martin Greenberger, John Korbel and Alice M. Rivlin, Microanalysis of Socioeconomic Systems: A Simulation Study (New York: Harper, 1961), pp. 233-56.

Steps Toward An Economic Model

This study will undertake to resolve three sets of problems which must be confronted in the development of any economic model of household food expenditure which includes socio-economic and demographic variables.

The first set of problems essentially are ones of specification of a structural model. A structural model specifies the dependent variable and the independent variables which affect it. At the present stage, we can specify this structural model only in a very general form:

$$Y = f(X_1, X_2, \ldots X_K, X_{K+1} \ldots X_n)$$

where Y is household food expenditure, X_1 is household income, X_2 to X_K are variables relating to the size and composition of the household and to the number of meals eaten at home, and where X_{K+1} to X_n are certain variables representing household socio-economic and demographic characteristics. Our first set of problems is to determine which household S-D characteristics affect food expenditure and should be included in the structural model. These problems are considered in Chapter II.

The second set of problems are ones of variable specification. The problem of measuring the S-D characteristics included in the structural model is one such problem. The patterns of effects of the significant S-D variables and the implications of these patterns for the selection of variables to represent these characteristics is another such problem. These two problems and related issues are considered in Chapter III.

The third set of problems which will be considered arise in the development of an economic model which relates the economic and S-D variables to food expenditure and which specifies the functional form of this relationship. We will consider some of the alternative models

which have been employed and develop a classification of these models on the basis of the assumptions made about the functional form of the effects of household S-D characteristics. These matters are dealt with in Chapter IV.

Scope

Principal emphasis will be given to the problems involved in the development of a micro-level model of household food expenditure, which employs cross-sectional survey data. Because of the magnitude of the problems involved, it was necessary to stop short of the specification of a complete economic model and the estimation of the parameters of such a model.

This study will employ observations selected from the nationwide sample of the U. S. Department of Agriculture's 1955 Household Food Consumption Survey. The observations employed represent both rural and urban households throughout the nation. The availability of this data helps to overcome many of the problems which have limited the usefulness of earlier estimates of the effects of household S-D characteristics.

Since this study is focussed on problems involved in the development of an economic model of food expenditures which includes variables representing household S-D characteristics it was not possible to consider certain other problems involved in the analysis of household expenditure data except insofar as they relate to this study. For this reason only limited attention is given to the effects of household age-sex composition on food expenditures. The Standard Meal Units variable was employed to deal with these effects. Its derivation is, however, merely an adjunct to the central purpose of this study. Other problems which were, of necessity, considered only briefly were the identification of the unit of investigation and the specification of the income variable.

CHAPTER II

Specifying The Structural Model

The problems confronted in determining which household socio-economic and demographic (S-D) characteristics should be included in a structural model explaining household food expenditure behavior will be considered in this chapter. Problems related to the development of empirical measures of these variables are discussed in Chapter III. Alternative economic models specifying the functional form of the relationship of the economic and S-D variables to household food expenditures are considered in Chapter IV.

Selection of S-D Variables for Inclusion in the Structural Model

As was pointed out in Chapter I, the first set of problems which we confront in developing an economic model relating economic and S-D variables to household food expenditure is the problem of specifying which S-D variables should be included in such a model. Our first problem thus is to specify the structural model explaining food expenditures. We can, at this stage, specify this model only in the general form:

$$Y = f(X_1, X_2 \ldots X_k, X_{k+1} \ldots X_n)$$

where Y is household food expenditures, X_1 is household income, X_2 to X_k are variables representing household size and composition and the number of meals served and where variables X_{k+1} to X_n represent the socio-economic and demographic characteristics of the household. Our principal problem in this chapter is to determine which household characteristics significantly affect expenditures and should be represented by X_{k+1} to X_n.

Theoretical Bases for Selection

In undertaking to determine which S-D variables should be singled out for inclusion in the structural model we are without guidance from economic theory itself. Classical economic theory has treated the influence of membership in particular S-D categories as one aspect of "tastes", and thus as given. Classical theory thus provides us no guidance for selecting the most relevant characteristics from the multiplicity of all S-D variables. We do, however, have some basis for selection of relevant S-D variables without resorting to the raw empiricism of successively examining each conceivable S-D variable. These bases are the theories of other disciplines concerned with the influence of socio-demographic characteristics on behavior and the research results of other disciplines as well as those of economics itself.

The theories of other disciplines give rise to certain hypotheses about the effect of S-D characteristics upon expenditure behavior. The S-D variables in these hypotheses are employed, in most economic studies, without any real examination of the theories underlying the hypotheses. These theories, in fact, are often only sub-theories or of only rather rudimentary form. These theories do, however, underlie the use of S-D variables in the study of consumer expenditure behavior whether they are made explicit or not. The influence of most S-D variables is thus usually part of a theoretical structure, although, most often this structure is not a part of economics itself.

Since the influence of S-D variables upon consumer expenditure is not part of any systematic economic theory, there are few criteria by which to judge whether all relevant S-D characteristics have been dealt with in any particular study. Individual S-D variables can be seized

upon for study on the basis of "common sense". However, any comprehensive treatment of the influence of S-D variables demands a more organized approach.

The need for an organized structure detailing the influences affecting household food expenditure can be met, at least in part, by a modified version of Warren Bilkey's classificatory system.[1] Bilkey, after his review of literature in the area, classified the influences affecting consumer expenditure behavior into four major categories: (1) biologically based needs, drives and wants, (2) socio-cultural influences, (3) institutional-availability circumstances and (4) "immediate" influences such as income, prices, habits and expectations. A system of classification of the influences affecting food expenditure behavior based on Bilkey's system is given in Table 2.1.

Empirical Bases for Selection

Some indication of which variables among those listed in Table 2.1 are likely to prove most important can be gained by examining the results of past studies of the effects of household S-D characteristics on expenditures for food.

The 1950 BLS Survey tabulations provided the observations employed in Jean Crockett's examination of the effects of family size, city size, race, age of head and occupation. Mrs. Crockett found the income-expenditure elasticities for food (excluding alcoholic beverages) for home consumption to be affected by region, race and the self-employment status of the head. The level of expenditure (regression constant term)

---

[1] Warren J. Bilkey, The Basic Relationships in Consumer Expenditure Behavior ("Harvard Studies in Marketing Farm Products," No. 4-H; Cambridge: Harvard University, 1951), pp. 1-32, 64-65.

Table 2.1 Characteristics Explaining Differences in Food Expenditure Behavior Between Households [a]

Biologically based needs and wants

1. Physiological-nutritional needs as affected by:
   (a) age and sex composition (adult-equivalent scales) (stage in family life cycle)
   (b) occupations
   (c) prevailing temperatures and climate (region)
2. Psychological needs and wants

Socio-cultural Influences

1. Regional eating patterns (region)
2. Eating patterns related to urbanization (urbanization)
3. Ethnic eating patterns (race)
4. Religious proscriptions and prescriptions
5. Influence of socio-economic status
   (Social class)
   (Occupation)
   (Education of husband, or wife)
6. Influence of family life cycle
   (Age of husband, or wife)
   (Stage in family life cycle)
   (Presence of children, age of oldest or youngest child)

Institutional-availability Circumstances

1. Supply situation
   a. Meals from home supplies
      (1) Market situation
      (2) Home production activities
         (a) Home raising of products
            (Urbanization)
            (Value of home production)
         (b) Home processing activities
            (Employment of homemaker)
            (Access to freezing facilities, etc.)
         (c) Home preparation activities
            (Employment of homemaker)
   b. Meals not from home supplies
      (Employment of homemaker)
      (Occupation - lunch not carried from home)
      (Presence of school-age children, eating school lunch)
2. Advertising
3. Credit and purchase arrangements

"Immediate" Influences

1. Income level
2. Prices
   (Region - inter-regional price differences)
   (Urbanization - intra-regional price differences)
3. Past commitments and experiences
4. Habits and current desires
5. Expectations and goals for future

---

[a] This table is based on the classification developed by Warren J. Bilkey, The Basic Relationships in Consumer Expenditure Behavior ("Harvard Studies in Marketing Farm Products," No. 4-H; Cambridge: Harvard University, 1951), pp. 28-32.

Items given in parentheses are proxy variables which are often employed to represent the influence of particular characteristics.

was found to differ by race and age of the head.[1]

The 1955 USDA Survey data tabulations served as the basis of Miss Marguerite Burk's studies of the effects of region and urbanization. Miss Burk concluded that the effects of both factors were highly significant after an examination of the Engel curves of rural farm, rural non-farm and urban households in each of four regions.[2]

A study of households in Lansing, Michigan, by Thomas Moss examined the effect of certain household characteristics on total expenditures for food at home and for various food categories. On the basis of analysis of variance tests, Moss concluded that per capita expenditures for food were affected by family size and income but that the effects of age of the housewife, her education and the occupation of the household head were not statistically significant.[3]

The effects of four household characteristics, family size, income, age of the meal planner and education of the meal planner, on expenditures for meat were studied by Charles Zwick in his analysis of data collected from Medford, Massachusetts households in 1952 and 1953.[4] Zwick found statistically significant differences in income elasticities between households with younger and older meal planners. Income elasticities were not affected significantly by differences in family size or educational level, however. Short-run price elasticities were

---

[1] Crockett, pp. 293-310.

[2] Marguerite C. Burk, Influences of Economic and Social Factors on U. S. Food Consumption, (Minneapolis: Burgess, 1961), pp. 53-69.

[3] Thomas N. Moss, "Some Relationships of Selected Socio-Economic Factors to Food Consumption and Expenditures, Lansing, Spring 1950," (unpublished Ph.D. dissertation, Department of Agricultural Economics, Michigan State College, 1952), p. 141.

[4] Zwick, pp. 451-56.

found to be affected only by the age of the meal planner. Income, education and family size were not found to affect price elasticity to a statistically significant extent.

Viewed as a group, these four studies suggest that urbanization, region, race, self-employment status and the age of the homemaker or household head are important S-D variables affecting household food expenditures. All are to be given special attention when the significance of the effects of S-D characteristics is evaluated.

The Use of Proxy Variables

The influences listed in Table 2.1 are not all easily observed or ascertained by questioning. It is almost impossible for a survey respondent or an interviewer to determine to what extent the respondent's consumption preferences have been influenced by regional tastes, for example. Yet it is easy enough to determine the region in which the respondent resides. In using region of residence to represent the influence of regional tastes we are employing it as a proxy variable.

Looking over Table 2.1 we can see that many of the factors influencing consumer demand and expenditure are customarily represented by proxy variables because of the difficulty in observing or quantifying the influence itself. Proxy variables are expected to explain much the same portion of the variance of expenditure as the variable they represent. From the examples in Table 2.1, it can seem that many proxy variables will not explain all of the variance in expenditure arising from the variable which they represent. Others, however, may represent the effects of several underlying variables.

The Problem of Underlying Variables

Certain of the proxy variables include the effects of several different influences. The variable "region" includes the effects of regional eating patterns, regional price differences and regional supply differences. The effect ascribed to "region" in any analysis of food expenditures is thus the net effect of these and other region-related influences.

The use of variables, such as region, which combine the effects of several influences, interferes with the examination of the separate effect of any particular influence. We often have little way to get at these underlying influences since we lack means to measure them; often we even lack words to describe them. For the present we must content ourselves with the analysis of the effects of variables which often include several different influences each of which we would like to analyze separately. We will, then, deal with the structure imposed on the analysis by the concepts used to collect the data, but should remind ourselves, from time to time, about the combination of influences which underlie them.

The S-D Variables Examined

Since little is known about the significance of the customarily employed S-D variables or the pattern of their effects, it appeared most fruitful to consider these basic questions before considering the problem of developing an economic model including several such variables.

We will consider the following variables:

Urbanization

Region

Race

Stage in Family Life Cycle

Social class

Education of homemaker

Self-employment of household head

Employment of homemaker

These variables are, in general, of the form customarily employed in surveys of household expenditure. As has been noted, certain categories such as region estimate the net effect of a number of different influences (temperature, regional food habits, regional price differences) related to region. Several of the S-D variables which will be considered are only proxy variables for the basic influences set forth in Table 2.1. The variables listed will be employed with full knowledge of these shortcomings in the hope that the results of this study will provide a basis for future improvements and refinements.

The variables to be considered include only a part of the influences listed by Bilkey. We have chosen to focus on only certain of these influences, those which can be regarded as arising from either the demographic characteristics of the household or social group membership. All such influences will not, however, be considered. Certain types of information which would have been of interest were not collected in the 1955 USDA Survey. Information on religious affiliation and ethnic origins (other than race) was not requested. Education of the husband or male head was not requested. We do however, have eight important variables for which we do have information. We will proceed with the analysis of the effects of these variables, with the knowledge that certain other socio-demographic variables have, of necessity, been omitted from the analysis.

Choice of Research Strategy

Before proceeding to examine individual S-D characteristics and test their significance, it was necessary first to determine the overall strategy of research to be employed. In considering the effects of many factors two approaches are possible, as James Morgan and John Sonquist have pointed out.[1] One procedure is to look at one factor at a time, keeping in mind the effects of other variables not included in the analysis. The second and more complex procedure is to consider all the factors regarded as relevant more or less simultaneously. This may be done, for example by including them all in a single multiple regression equation. Such a procedure requires the immediate imposition of a number of restrictions on the forms of the variables and may necessitate the use of scales and indexes which are often developed on the basis of rather arbitrary decisions. The development of the dummy variable technique of representing qualitative variables has eased this problem considerably, Morgan notes.

It was felt that a separate preliminary investigation of the effects of each variable was necessary before combining them all in a single multi-variate analysis. Such an approach permits a careful formulation of the categories of any particular S-D characteristic and evaluation of the relative merits of two variables which can be hypothesized to explain the same portion of the variance of expenditure. The examination of one S-D variable at a time also avoids the problems of multi-collinearity which could be expected to arise in a regression analysis including a large number of socio-demographic characteristics. The presence of

---

[1] James N. Morgan and John A. Sonquist, "Problems in the Analysis of Survey Data, and a Proposal," Journal of the American Statistical Association, Vol. 58 (June, 1963), p. 421-22.

multi-collinearity would both increase the standard errors of the coefficients estimated and seriously interfere with the tests of significance planned. Theil notes that variables which are intercorrelated with other independent variables are likely to be judged non-significant and that this may lead to the rejection of variables which properly should be included in a model.[1] Theil adds the observation that when several variables are employed the intercorrelations do not need to be large to produce misleading non-significant results.

The examination of a single variable at a time does give rise to one serious problem, however. Hanan Selvin has pointed out that researchers using survey data to determine the effects of a particular characteristic on individual behavior often fail to hold constant correlated factors which also influence behavior.[2] W. G. Cochran has referred to the effects of these correlated factors as "correlated biases".[3] It was felt that the inclusion of an income variable and a variable taking account of both the number of meals eaten by members of the family and the family composition would remove three important sources of such bias. After the initial tests of the S-D variables, certain variables which were suspected to be influenced by correlated biases were subjected to further tests. In these tests, the variables representing potential sources of bias were also included in the analysis. In this way, the significance of the variable under study could be determined free from the effects of the principal sources of correlated biases.

---

[1] H. Theil, Economic Forecasts and Policy. ("Contributions to Economic Analysis," No. 15; Amsterdam: North-Holland Publishing Co., 1961), pp. 355-56, 216-17.

[2] Hanan C. Selvin, "A Critique of Tests of Significance in Survey Research," American Sociological Review, Vol. 22 (October, 1957), pp. 521-22.

[3] W. G. Cochran, Sampling Techniques, (New York: Wiley, 1953), p. 305.

The decision to employ a single-equation model and rely principally on the test of one S-D characteristic at a time involved the acceptance of several restrictive assumptions. It was felt, however, that the research strategy chosen allowed maximum flexibility and avoided a premature selection of a more elaborate model involving a far larger number of assumptions. The selection of the preliminary economic model which was employed and the assumptions involved in its use are discussed in the next section.

Selection of a Preliminary Economic Model

Examination of Alternative Regression Models

A number of alternative functional forms have been employed in regression analysis for the expression of the relationship of income to food expenditure. In previous work with data from the U. S. Department of Agriculture 1955 Survey the author found two forms which seemed to be superior to the others which might be employed.[1] The two models were the semi-logarithmic equation form:

$$Y = a + b \log X_1 + u \qquad (2.1)$$

and the double-logarithmic form:

$$\log Y = a + b \log X_1 + u \qquad (2.2)$$

where Y is one week's expenditure in Spring 1955 for food and beverages consumed at home or taken from home supplies for consumption elsewhere,

---

[1] Robert O. Herrmann, "An Investigation of Differences in Income Elasticities of Demand for Food in Households of Differing Size and Composition," (unpublished Master's thesis, Department of Agricultural Economics, Michigan State University, 1961), p. 104.

X_1 is 1954 income after Federal and State income taxes and the u's are independently distributed with mean zero, variance σ^2. Since both models had performed about equally well, on the basis of comparisons of estimated adjusted coefficients of multiple determination, it was decided to compare their performance after adjustment of the data for differences in family size.

In order to be able to estimate the parameters of the relationship of income to food expenditure it is necessary to hold other factors which affect the level of expenditure constant between households. The other factors which differ between households and directly affect the level of expenditure include the size and composition of the family, the number of meals eaten at home and prices paid. Since a single Engel curve for all sizes of families was to be estimated, it was necessary to devise some method to take account of the factors, other than income, which directly affect the level of expenditures.

A commonly used technique of adjusting food expenditures for differences in family size is to place them on a per capita or per adult-equivalent basis.[1] The per capita adjustment involves simply the division of total expenditure by the number of members in the family. The adjustment to expenditure per adult equivalent employs age-sex specific scales, in which the consumption of an adult male is usually assigned the value 1.0, with infants, children and so forth, assigned smaller values describing their relative levels of consumption. In this

---

[1] Several alternative techniques of adjusting for differences in household size and their performance are discussed in Faith Clark, Janet Murray, Gertrude S. Weiss and Evelyn Grossman, Food Consumption of Urban Families in the United States with an Appraisal of Methods of Analysis, U. S. Dept. of Agriculture, Agriculture Information Bulletin No. 132 (Washington: USGPO, 1954), pp. 35-39.

adjustment the sum of the scale coefficients of the family replaces the number of family members in the denominator when dividing total expenditure. The direct adjustment of expenditure to a per capita or per adult equivalent basis has the disadvantage that it does not take account of economies of scale in household operation separately from the relationship of income and expenditure. Since work by Prais and Houthakker and by Janet Murray has indicated that such economies do, in fact, exist, the failure to take them into full account must be regarded as a serious shortcoming.[1]

An alternative method of adjustment suggests itself. This is the use of the number of family members or the sum of scale coefficients as an additional independent variable. Such a formulation separates the relationship of family size and food expenditure from the usual curvilinear relationship of income and food expenditure. Appropriate transformations of the family size variable can be employed to permit a curvilinear relationship of family size and expenditure if economies of scale do exist.

In addition to the effects of differences in family size and composition, total expenditures are affected by differences between households in the number of meals which are eaten at home. The total number of meals eaten at home depends both on the size of the family and upon the extent to which individual members utilize alternative food sources. Previous work with the data indicates that substantial variation in the number of meals eaten at home exists even within households of a given

---

[1] S. J. Prais and H. S. Houthakker, The Analysis of Family Budgets ("University of Cambridge, Department of Applied Economics Monographs," No. 4; Cambridge: Cambridge University Press, 1955)., pp. 146-52 and Janet Murray, "Per Person Food Cost Differentials in Large and Small Families," U.S. Department of Agriculture, Family Economics Review (September, 1960), pp. 3-4.

size.[1] For this reason it was considered important that the effects of both family size and composition and of number of meals eaten at home on total expenditures be taken into consideration.

The effects of differences in the total number of meals eaten at home have been adjusted in some U.S.D.A. studies through the use of an "equivalent persons" measure. Despite its name this term relates to the number of meals served rather than the size or composition of the family. The total number of meals served to members in the household from family food supplies is divided by 21 to obtain the household size in "equivalent persons".[2] The adjustment of food expenditure to a per equivalent-person basis involves the same disadvantages as adjustment to a per capita or per adult-equivalent basis, in that no separate account is taken of possible economies of scale. This suggests that either total meals served or household size in equivalent persons should be employed as an independent variable, with such transformation as may be suitable.

It was decided to combine the adjustments for size-composition differences and differences in total meals served into a single independent variable. A variable taking account of both family size and composition and of the total number of meals was then devised. This variable was labeled the Standard Meal Units (SMU) variable.

The variable was obtained by the following method:

$$T \frac{\sum_j \sum_i k_{ij}}{\sum_i n_i} = SMU \qquad (2.3)$$

$$i = 1, \ldots \qquad j = 0, 1, \ldots J$$

---

[1] Herrmann, p. 71.

[2] U.S. Department of Agriculture, Food Consumption of Households in the United States, Household Food Consumption Survey 1955 Report No. 1 (Washington: USGPO, 1956), p. 194.

where T is the total number of meals served members of the family in the survey week, where the K_{ij} are the age-sex specific scale coefficients of the I different age-sex categories into each of which j members of a particular household are classified, and where n_i is the number family members in category i. The SMU variable is thus the total number of meals served standardized for differences between families in the age-sex composition of the individuals served. The derivation of the SMU variable is discussed in detail in Appendix II.

Prices paid were assumed to be the same for all households in this study. This source of differences in expenditures between households will not be considered specifically in this study. Since the data were collected in cross-section, the assumption of no differences in prices paid is probably a valid approximation. Certain differences do exist, such as regional price differences and will, of course, be part of the effects associated with "region" of residence.

Three alternative regression models were examined in order to determine (1) the proper expression of the relationship of income and expenditure, (2) the usefulness of the SMU variable, and (3) the proper expression of the SMU variable, if the use of such a variable appeared desirable.

The three models and the results obtained are given in Table 2.2. An examination of the $\overline{R^2}$ coefficients obtained with the deletion of the SMU variable (X_2) for equations 2.4 and 2.6 indicates that the double-logarithmic expression of the relationship of income and expenditure (equation 2.2) yields a higher estimated adjusted coefficient of

Table 2.2 Alternative Models Examined in Selecting a Preliminary Economic Model

	Model	Regression Constant Term a	Regression Coefficient b	Regression Coefficient c	$\bar{R}^2$	$\bar{R}^2$ with deletion of X_2 or X_3
(2.4)	$Y = a + b \log X_1 + c X_2 + u$	-52.9949 (2.10203)	18.32563 (.59517)	.17949 (.00553)	.40	.23
(2.5)	$Y = a + b \log X_1 + c \log X_2 + u$	-78.44913 (2.19469)	17.07871 (.59561)	23.59870 (.69495)	.41	.23
(2.6)	$\log Y = a + b \log X_1 + c \log X_2 + u$	-.06513 (.04155)	.37306 (.01128)	.53278 (.01316)	.50	.27
(2.7)	$Y = a + b \log X_1 + d X_3 + u$	-54.55811 (2.10314)	18.32114 (.59323)	.14443 (.00447)	.40	.23
(2.8)	$Y = a + b \log X_1 + d \log X_3 + u$	-81.83854 (2.22674)	17.66341 (.59151)	23.21298 (.68191)	.41	.23

$n = 3641$

$\bar{Y} = \$22.38$ $\quad \overline{\log Y} = 1.29551$

$\bar{X}_1 = \$4505.29$

$\bar{X}_2 = 57.70$ $\quad \overline{\log X_1} = 3.57522$

$\bar{X}_3 = 70.24$ $\quad \overline{\log X_2} = 1.70625$

$\overline{\log X_3} = 1.78979$

multiple determination than the semi-logarithmic expression.[1] Since there was no reason to prefer one model over the other, the double-log model was selected on the basis of its ability to explain a higher proportion of the variance in food expenditures.

A comparison of the estimates of $\bar{R}^2$ obtained with the inclusion of the SMU variable (X_2) with estimates of $\bar{R}^2$ with the deletion of X_2 indicates that the SMU variable explains an important portion of the total variance in food expenditures. The addition of the SMU variable increased $\bar{R}^2$ from .27 to .50 in the double-log model (2.6) selected.

It was hypothesized that a formulation permitting economies of scale for the SMU variable would yield superior results to one which did not. A formulation in which the change in expenditure resulting from an increase in size (expressed in SMU's) is inversely proportional to SMU permits economies of scale, if they are present. We can see that equation (2.5) permits economies of scale, while (2.4) does not. This is because the change in expenditure with a change in SMU is constant in (2.4), i.e.

$$\frac{dY}{dX_2} = c$$

while it is inversely proportional to X_2 in equation (2.5), since for

---

[1] The adjusted coefficient of multiple correlation ($\bar{R}^2$) has been employed throughout this study. $\bar{R}^2$ is an adjustment of the statistic R^2 to take account of the tendency of R^2 to overestimate the proportion of the variance of the dependent variable explained when the number of parameters estimated is large or the number of observations is small. The relation of $\bar{R}^2$ to R^2 is:

$$\bar{R}^2 = 1 - (1-R^2)\ \frac{n-1}{n-k}$$

where n is the number of observations and k the number of independent variables. See Earl O. Heady and John L. Dillon, Agricultural Production Functions (Ames: Iowa State University Press, 1961), pp. 118-19.

equation (2.5)

$$\frac{\partial Y}{\partial X_2} = \frac{c}{X_2}$$

A comparison of the results for equations (2.4) and (2.5) indicates that the use of a logarithmic transformation did improve the $\bar{R}^2$ obtained slightly. SMU, therefore, was inserted in the double-log formulation with a logarithmic transformation. This formulation (2.6) also permits economies of scale since the partial derivative of expenditure with respect to SMU is inversely related to size in SMU, i.e.,

$$\frac{\partial Y}{\partial X_2} = \frac{cY}{X_2}$$

The development of the SMU variable (X_2) required a number of additional steps in the analytical process. These steps were both time-consuming and costly in terms of personnel and equipment operation expenses. For this reason it seemed worthwhile to ascertain whether the SMU variable performed better than the use of a total meals served variable without standardization. For the purposes of comparison two additional equations (2.7) and (2.8) were estimated in which X_3, total meals served members of the family, replaced X_2, standard meal units served. A comparison of the results for (2.7) with those for (2.4) does not indicate that the use of SMU, rather than total meals served, produces any improvement in the proportion of the variance in expenditures explained. A comparison of the results for (2.8) with those for (2.5) suggests the same conclusion. It was decided, however, to employ the SMU variable in this study, rather than total meals served. This decision was made on the basis of the a priori preference for the SMU formulation. It appears that when differences in the size of the family

and the proportion of meals eaten at home are taken into account through the use of the total meals served variable, further adjustment for differences in family composition provides little improvement in the $\overline{R^2}$ obtained.

The simple correlation between log income and log SMU was found to be .23. Because of this positive correlation the inclusion of the SMU variable reduced the size of the coefficient of the income variable. It will be noted that the constant term and the coefficient of the income variable are larger in the equations employing the total meals served variable than in the comparable ones employing the SMU variable. This difference is due, at least in part, to the fact that the correlation of income and total meals served is slightly lower than the correlation of income and SMU. The simple correlation of log total meals served and log income was found to be .20.

After review of the results noted above, a preliminary economic model of the form of equation (2.6) was selected:

$$\log Y = a + b \log X_1 + c \log X_2 + u \qquad (2.6)$$

This model provides the double-logarithmic expression of the relationship of income to expenditure and also provides for the use of a logarithmic transformation of the SMU variable which permits economies of scale, if they are present.

The Insertion of the Socio-demographic Variables in the Preliminary Model

It was decided to specify the socio-demographic characteristics under consideration as dummy variables, in an equation of the form of equation (2.6). The equations employed are thus of the form:

$$\log Y_t = a + b \log X_{1t} + c \log X_{2t} + \sum_{i=2}^{I} d_i Z_{it} + u_t \qquad (2.9)$$

$$t = 1, \ldots n$$

where Z_{it} takes the value 1 when characteristic i is present for household t and 0 when it is not, and I is the total number of categories of a given socio-demographic characteristic. When the constant term is employed, dummy variables for all I categories of the characteristic cannot be entered into the equation. If dummy variables were entered for all I categories, then $\sum_{i=1}^{I} d_i Z_{it}$ would be linearly dependent on the variable representing the constant term. For this reason, dummy variables for only I-1 of the I categories of a S-D characteristic are employed.[1]

Several considerations prompted the use of dummy variables to represent the effects of the S-D variables under consideration. With the use of dummy variables the effects of I-1 categories of any given S-D characteristic can be represented. The model is simple and easily interpreted. The representation of the effects of the different categories in a single equation gives the advantage of both simplicity and economy. The number of different equations which must be estimated is held to a minimum.

The use of dummy variables permits us to ascertain the separate effects of membership in each different category. The use of dummy variables thus does not assume that the effects of moving from any one category to the next are equal or that these effects are linear as is assumed when a continuous variable is used to represent a S-D characteristic. With the use of dummy variables we are able to identify any patterns in the effects of the S-D characteristic on expenditure which may exist, since the use of dummy variables involves only the assumption that the effects

---

[1] See Daniel B. Suits, "Use of Dummy Variables in Regression Equations," Journal of the American Statistical Association Vol. 52 (December, 1957), p. 549.

of the individual categories are additive constants. In the present case, since we are using the double-logarithmic model the assumption is modified slightly. When dummy variables are used with this model it as assumed that membership in a given category, rather than the omitted category, affects expenditure by a constant percentage amount.

The use of dummy variables does have the disadvantage that the effects of the omitted category cannot be determined. Since the principal emphasis in this portion of this study is on the significance of the effects of S-D variables, rather than on measuring the magnitude of these effects, this consideration was of little consequence.

The Tests of Significance Employed

The Statistical Test Employed

Covariance analysis employing the model (2.9) has been used as the basis for the test of the significance of individual S-D characteristics. The test employed is an F test of the null hypothesis that the coefficients of the Z_i in (2.9) are equal to zero. The test procedure will be discussed in detail in the remainder of this section.

It was desired to test the significance of each of the S-D characteristics under consideration, each consisting of I categories. This required the testing of the significance of the categories of the characteristic as a group. The model selected:

$$\log Y = a + b \log X_1 + c \log X_2 + \sum_{i=2}^{I} d_i Z_i + u \qquad (2.9)$$

since it employs an overall constant term, a, has omitted the first of the I categories. (Category 1 of the I categories will be regarded as the omitted category). As Gustafson has pointed out, the d_i ($i = 2, \ldots I$) coefficients are an estimate of the effect on the dependent variable

of a change in membership from the omitted category to another category.[1] Stated alternatively, the d_i are estimates of the difference between the constant term of the omitted category, (a_1) and the constant term of category i, a_i $(i = 2 \ldots I)$, i.e., $d_i = (a_i - a_1)$, where a_1 and a_i are estimated by:

$$\log Y = a_1 + b_1 \log X_1 + c_1 \log X_2 + u \qquad (2.10)$$

$$\begin{matrix} \log Y = a_2 + b_2 \log X_1 + c_2 \log X_2 + u \\ \vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \\ \log Y = a_I + b_I \log X_1 + c_I \log X_2 + u \end{matrix}$$

A test of the hypothesis that all of the categories, taken jointly, do not significantly affect the dependent variable is equivalent to a test that all the a_i $(i = 1, \ldots I)$ are equal, or that the d_i are equal to zero, i.e.,

$$a_i - a_1 = d_i = 0 \qquad (i = 2, \ldots I) \qquad (2.11)$$

The F test employs the sum of squares due to regression (SSR) and the sum of squared residuals (SSE) under the alternative hypothesis, which is an equation of form (2.9). It also employs the SSR under the null hypothesis that the d_i are equal to zero. The null hypothesis is equivalent to (2.11), and the SSR under the null hypothesis, SSR_{Ho}, was estimated by computing an equation of the form of (2.9) which omitted the Z_i, the dummy variables. Under the null hypothesis the test statistic

$$\frac{SSR_{Ha} - SSR_{Ho}}{SSE_{Ha}} \cdot \frac{n-r-1}{r-k}$$

---

[1] Robert L. Gustafson, "The Use and Interpretation of 'Dummy Variables' in Regressions," Note of January 22, 1962 (Revised), Department of Agricultural Economics, Michigan State University, East Lansing (Mimeographed), pp. 1-4.

has an F distribution with (r-k,n-r-1) degrees of freedom, where n is the number of observations, k the number of independent variables under the null hypothesis and r is the number of independent variables under the alternative hypothesis.[1]

The category selected for omission in estimating the regressions of form (2.9) for the tests of significance was usually the category whose behavior is considered standard or average. For example, in the regression including dummy variables for region, the omitted category was "North Central". It was felt that the differences between this category and each of the other three would be of the greatest interest. A similar pattern was followed in selecting the category omitted for each of the other characteristics. The statistical tests are not affected by the choice of omitted category.

One of the important advantages of covariance analysis with the regression model is that it permits tests of the significance of the effects of the categories of a given characteristic and at the same time provides estimates of the differences between categories in these effects. The dual nature of these results contrasts with those which can be obtained with certain analysis of variance models. The differences in the results provided by alternative models of the effects of S-D variables will be discussed in detail in Chapter IV.

The significance of the differences in the effects of the categories of the S-D characteristics under examination will be considered in the remainder of this chapter. The pattern of these differences in effects, the d_i estimated and the implications of these patterns for the specification of S-D variables will be considered in the following chapter, Chapter III.

---

[1] R. L. Anderson and T. A. Bancroft, Statistical Theory in Research (New York: McGraw-Hill, 1952), p. 172.

A Comment on the Effect of Sampling Procedure on the Tests of Significance and on Estimates of Confidence Intervals

The sample from which the household observations employed in this study were drawn was a complex stratified self-weighting probability sample. The sampling technique employed thus differed from the simple random sampling technique which is assumed to underlie most of the familiar statistical techniques, including regression analysis. The sampling procedures of most survey research, including the 1955 USDA Survey, do not conform to simple random sampling procedures.

Leslie Kish has pointed out that clustered sampling procedures, i.e., questioning two respondents in the same household, the same neighborhood or the same job at work, can produce substantial intra-class correlations. The effect of these correlations is to reduce variation in the sample, causing the variance estimate to underestimate the true population variance.[1]

Kish's empirical investigations of the problem indicate that the effects of clustered sampling were least serious in the sample whose design most resembled the one employed in the 1955 USDA Survey.[2] It is expected, therefore, that the effects of clustered sampling procedures are not too serious. The F-ratios, standard errors and other statistics presented in this study have been estimated with procedures which assume random sampling techniques.

In general, the clustered sampling may cause the standard errors calculated to underestimate the true standard errors and may cause the null hypothesis to be rejected too often in tests of hypotheses. These

---

[1] Leslie Kish, "Confidence Intervals in Clustered Samples," American Sociological Review, Vol. 22 (April, 1957), pp. 154-65.

[2] Ibid., p. 158.

possible effects should be borne in mind in examining the results which will be presented.

Test of Economic Significance

In addition to the test of statistical significance, a test of economic significance will be applied. It was felt that if the differences in expenditure between categories were small they would be of little economic significance even if they should prove to be significant in a statistical sense.

The economic test criterion provides that a characteristic shall be judged economically significant if the difference between the mean expenditure of all households and those of households in any one category of a characteristic exceeds one dollar.[1] The dollar level chosen for the test is slightly more than five percent of the geometric mean of expenditures for all households included in the study, $13.74. Thus a characteristic will be adjudged significant from an economic standpoint if there is a difference in expenditure between one category of households and the overall mean in excess of five percent (approximately) of the overall mean.

Analysis and Results

The tests of significance of the individual S-D characteristics will be reported in the following section.

---

[1] The means employed for this test are geometric means of expenditures rather than an arithmetic means.

$$\text{Geometric mean} = \text{Anti-log} \frac{\sum_{i=1}^{n} \log Y_i}{n}$$

See Frederick E. Croxton and Dudley J. Cowden, Applied General Statistics, (2nd ed.; Englewood Cliffs, New Jersey: Prentice-Hall, 1956), pp. 198-200.

Urbanization

In her study of economic and social factors influencing U.S. food consumption, Marguerite Burk identified urbanization as a key factor influencing food expenditures.[1] It therefore was expected that urbanization would prove to be both statistically and economically significant. As Miss Burk has pointed out, the net effects of urbanization are a combination of many influences: home production activities, the accessibility of certain types of food stores, and economic and sociocultural factors which influence expenditure but whose distribution differs by urbanization.[2]

The three urbanization categories employed were urban, rural nonfarm and rural farm. Households were assigned to categories on the basis of the Census of Agriculture definitions in use at the time of the 1955 U.S.D.A. Survey. Households were assigned to categories on the basis of the following definitions:[3]

> "Urban households lived in communities of 2,500 or more persons or in the fringe areas around cities of 50,000 or more."
>
> "Farm households were those that included a farm operator, a person responsible for the operation of a farm, either performing the labor himself or directly supervising it. A farm was defined as in the U.S. Census of Agriculture, i.e. a place of 3 or more acres with value of farm products raised (for sale or home use exclusive of home gardens) amounting to $150 or more in 1954 or a place of less than 3 acres with value of sales of agricultural products amounting to $150 or more."

---

[1] Marguerite C. Burk, Influences of Economic and Social Factors on U.S. Food Consumption (Minneapolis: Burgess, 1961), p. 53-4.

[2] Ibid., p. 54

[3] U.S. Department of Agriculture, Food Consumption of Households in the United States, p. 195.

> "Rural nonfarm households were those living outside of urban places that were not classified as rural farm."

A few farm households that lived in urban areas were assigned to the urban category.

The urban category was selected as the omitted category.

The urbanization characteristic was found to be significant at the five percent level. A difference of $7.07 was estimated to exist between the geometric mean expenditure of all households and those of rural farm households at the geometric mean values of income and SMU. On this basis of the results of these tests, the urbanization characteristic appears to be highly significant in both a statistical and an economic sense.

The details of the F tests are presented in Table 2.3.

Region

In addition to urbanization, Marguerite Burk has focussed on region as a second key S-D variable influencing food expenditure.[1] It was expected that it, too, would prove to be highly significant in both a statistical and an economic sense.

The four regions employed in the analysis were Northeast, North Central, South and West. States were assigned to their respective regional categories on the basis of the classification employed in the Census of Population. States were assigned to regions as follows:[2]

---

[1] Burk, p. 53.

[2] U.S. Department of Agriculture, Food Consumption of Households in the United States, p. 195.

Table 2.3. Tests of Statistical Significance

Characteristic Under Test	n	SSR_{Ha}	SSE_{Ha}	SSR_{Ho}	Test Statistic	$F_{.05}$	Computed F
Urbanization	3641	124.33	100.37	111.22	$F_{.05}(2,3636)$	2.99	237.43**
Region	3641	114.67	110.02	111.22	$F_{.05}(3,3635)$	2.60	38.05**
Race	3596	108.84	111.87	108.62	$F_{.05}(1,3592)$	3.84	7.18**
Life Cycle	3641	113.72	110.98	111.22	$F_{.05}(10,3628)$	1.83	8.16**
Social Class	3641	112.78	111.91	111.22	$F_{.05}(6,3632)$	2.09	8.41**
Education of the homemaker	3469	104.48	105.78	102.45	$F_{.05}(5,3641)$	2.21	13.29**
Self-employment of the head	3641	114.35	110.35	111.22	$F_{.05}(1,3637)$	3.84	103.29**
Employment of the homemaker	3513	103.67	109.08	103.57	$F_{.05}(1,3509)$	3.84	3.16

** Significant at the 1 percent level or less.

Northeast

Connecticut	New Hampshire	Pennsylvania
Maine	New Jersey	Rhode Island
Massachusetts	New York	Vermont

North Central

Illinois	Michigan	North Dakota
Indiana	Minnesota	Ohio
Iowa	Missouri	South Dakota
Kansas	Nebraska	Wisconsin

South

Alabama	Kentucky	South Carolina
Arkansas	Louisiana	Tennessee
Delaware	Maryland	Texas
District of Columbia	Mississippi	Virginia
Florida	North Carolina	West Virginia
Georgia	Oklahoma	

West

Arizona	Montana	Utah
California	Nevada	Washington
Colorado	New Mexico	Wyoming
Idaho	Oregon	

The North Central region was selected as the omitted category.

The regionality characteristic was found to be significant at the five percent level. A difference of $1.96 was found between the geometric mean expenditure of all households and that of households in the Northeast at the geometric mean values of income and SMU. The regionality characteristic was judged to be very significant in both a statistical and an economic sense.

Race

Differences in negro and white food expenditures have been reported in several past consumer surveys. Differences between negro and white households were found in the New York City portion of the 1935-36 B.L.S. study of family expenditures. Low-income negro families were

found to spend less for food than white families at comparable income levels.[1] The importance of such differences as were found is somewhat difficult to determine since the results are complicated by the employment of many negroes in domestic and restaurant jobs in which meals are part of the wage. More recently, a 1947-48 study of food expenditures by negro and white families in Richmond and Washington found the differences in expenditure to be greatest at the lowest income level.[2] The 1950 B.L.S. study of the expenditures of urban households reported that food expenditures were slightly larger in white households than in negro households at the same level of income.[3] The difference was greatest at the lower income levels. Because of the differences in food expenditure found between negro and white households in these past studies it was hypothesized that similar differences might be found on examination of the 1955 USDA survey data.

Two racial categories were employed: White and Non-white. The non-white category consisted almost entirely of negro households. The white category was the omitted variable.

The race characteristic was significant at the five percent level. A difference of $1.23 was found between the geometric mean expenditure of all households and non-white households, at the geometric means of

---

[1] U.S. Bureau of Labor Statistics, Family Income and Expenditure in New York City, 1935-36, Study of Consumer Purchases, Urban Series, Bulletin 643, Vol. II, p. 98 as cited in Willard W. Cochrane and Carolyn Shaw Bell, The Economics of Consumption (New York: McGraw-Hill, 1956), pp. 199-201.

[2] Helen M. Humes, "Family Food Expenditures, 1947 and 1948," U.S. Department of Labor, Monthly Labor Review, Vol. 66 (June, 1949), pp. 621-30.

[3] University of Pennsylvania, Study of Consumer Expenditures, Incomes and Savings, Vol. III: Expenditures for Food, Beverages and Tobacco, pp. 138-40.

income and SMU. On the basis of these tests the race characteristic appears to be a rather significant one. It was felt, however, that the results were likely to have been affected by other factors correlated with race. It was suspected that regionality effects, in particular, might have influenced the test. The effects of region and urbanization on the results will be examined later in this chapter.

Stage in the Family Life Cycle

In their study of consumer finances over the family life cycle, John B. Lansing and James N. Morgan found substantial differences between life cycle stages in incomes, rates of durable goods acquisition and size of savings holdings.[1] Other studies have found similar differences as well as differences in types of goods purchased, degree of preference for nationally advertised brands and awareness of advertising messages.[2] On the basis of these differences there was reason to expect differences between life cycle stages in food expenditure behavior. There is some evidence of such differences in previous studies of food expenditures.

Age of the homemaker, a concept closely related to family life cycle, has been employed in several food expenditure studies. Age of the family meal planner was found to influence consumer expenditure behavior for meat by Charles Zwick in his study of Medford, Massachusetts households.[3]

---

[1] John B. Lansing and James N. Morgan, "Consumer Finances over the Family Life Cycle," Consumer Behavior, Vol II: The Life Cycle and Consumer Behavior, Lincoln H. Clark (ed.), (New York: New York University Press, 1955), pp. 36-51.

[2] S. G. Barton, "The Life Cycle and Buying Patterns," pp. 53-57 and Donald L. Miller, "The Life Cycle and the Impact of Advertising", pp. 61-65, Consumer Behavior, Vol II: The Life Cycle and Consumer Behavior.

[3] Zwick, pp. 451-56.

Zwick found income elasticities for meat to be lower in households whose meal planner was under 40 than in those whose meal planner was over 40. Zwick laid the difference to the heavy expenditures of younger families for durable goods. Zwick also found price elasticities for meat to be lower in the older households than in the younger households.

Food purchases per person were found to first increase with age of homemaker then decline in each of four income categories in cross-classifications of the 1955 USDA survey data.[1] A similar parabolic relationship between age of the household head and level of total household food expenditure was found by Jean Crockett in her analysis of data from the 1950 B.L.S. Survey.[2]

In undertaking to study the effect of stage in the family life cycle on food expenditure behavior, we find a number of alternative formulations of the life cycle concept. Lansing and Morgan suggest several alternative sets of life cycle categories.[3] The categories suggested are based on a single-track conception of family life, with most households moving through each successive stage. In the formulation of life cycle categories for this study, a multi-track life cycle system seemed indicated. Such a system does not assume that the typical individual or couple will pass through each successive stage, but assumes only that they will move between categories. The system of categories developed for

---

[1] U.S. Department of Agriculture, Food Consumption and Dietary Levels as Related to Age of Homemaker, United States by Region, "Household Food Consumption Survey 1955," Report No. 14 (Washington: USGPO, 1959), p. 12.

[2] Crockett, pp. 306-309.

[3] John B. Lansing and Leslie Kish, "Family Life Cycle as an Independent Variable," American Sociological Review, Vol. 22 (October, 1957), pp. 512-19.

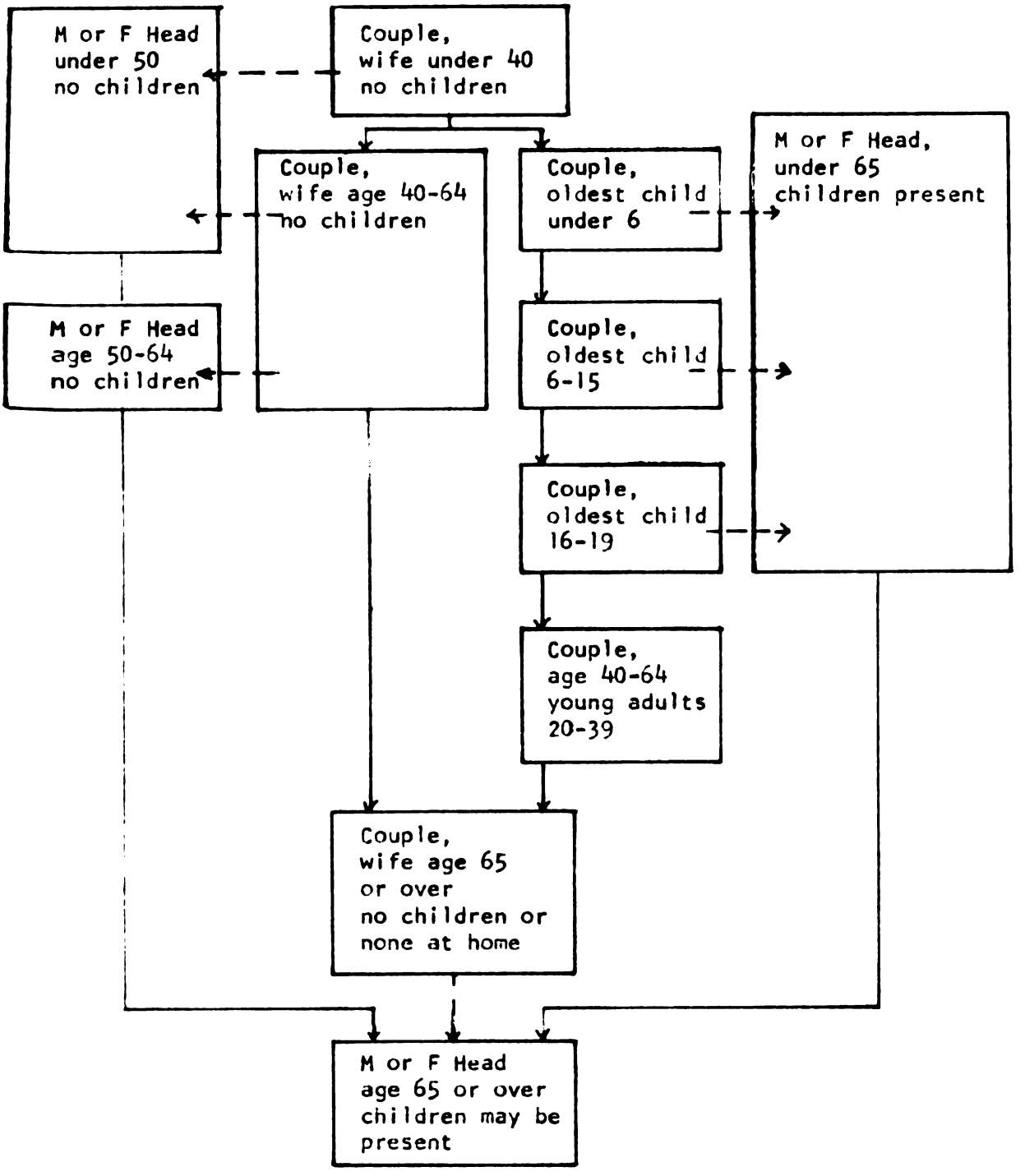
this study is given in Figure 2.1.

The arrows of Figure 2.1 indicate movement between stages. The solid line arrows indicate changes occurring chiefly with age or arrival and departure of children. The broken line arrows indicate changes occurring as the result of dissolution of a marriage by death of a spouse or divorce. Each stage has three dimensions or aspects: (1) marital status, (2) age (usually age of wife, or of oldest child) and (3) presence of children. Change in any one of these three dimensions moves a family into another stage category.

The importance of the marital status dimension is rather obvious. There is substantial reason to expect that the economic behavior of husband-wife households will be different from that of households headed by a male or a female, rather than a couple.

Age of the homemaker has been used as a kind of life cycle variable in several survey studies, as was mentioned. This concept was used in this study as one dimension of stage in family life cycle, rather than by itself. There is some reason to question why the use of the wife's age is to be preferred to the use of the husband's age. It is felt that the wife's biological capacity to bear children has a more important effect on the family life cycle than does the husband's age. Use of husband's age and focus on such points as probable age of retirement seems to inject economic considerations into what is essentially a sociological variable. Age of the children is another important age variable. In this study, age of the oldest child is used since it was felt that this formulation provided some indication of the time elapsed since the formation of the family by marriage. Certain other studies have employed the age of the youngest child. This formulation was employed by Lansing and Morgan, who considered it to be preferable to

Figure 2.1 The Multi-Track System of Family Life Cycle Stages Employed



other forms, but do not explain why.[1]

Presence of children was considered to be a third essential dimension of stage in the family life cycle. There is reason to believe that food expenditure habits are affected by the differences in the organization of the household and its consumption which are related to the presence of children. It should be noted that the compositional effects of the presence of children are taken into account in the Standard Meal Units variable. It is the effect of the presence of children on the organization and operation of the household of which we are speaking here. The stage in family life cycle variable is not designed to take account of differences in physiological requirements. The principal effects of compositional differences and physiological requirements are incorporated in the SMU variable.

One of the reasons underlying the development of the multi-track system of stages in the family life cycle employed was the desire to be able to classify all the households included in the study. A system was necessary which would provide for all types of households, some of which were not families in the strictest sense. Such a system had to include categories for "deviant" households in addition to the stages in the family life cycle usually emphasized.

The difficulty of expanding the family life cycle concept to include all households illustrates some of the problems of adapting theoretical concepts to applied research. A strict application of the family life cycle concept without expanding it to include "deviant" households would have necessitated discarding all the survey observations of "deviant" household types. The expansion of the concept permitted the use of these households in this study.

---

[1] Lansing and Morgan, p. 37.

Eleven stage categories were employed. The definitions for each category are given in Figure 2.1. The omitted category was "couple, with oldest child 6 to 15."

The stage in the family life cycle characteristic was found to be significant statistically at the five percent level. The category whose estimated expenditure was most different from the mean expenditure was the "Male or Female head over age 65" category. The estimated geometric mean expenditure of households in this category was $3.59 less than the geometric mean expenditure of all households, at the geometric mean values of income and SMU.

Social Class

The effects of social class on consumer behavior apparently have been neglected by economists because of their feeling that social class membership is so highly correlated with income that the additional variable has little explanatory power. A study by Joseph Kahl and James Davis suggests that income and social class membership may not be so highly correlated as is believed generally. In a study in Cambridge, Massachusetts Kahl and Davis found a tetrachoric correlation of .56 between income and the interviewer's rating of survey respondents' social class membership.[1] In explaining the rather weak relationship between income and social class membership, the authors note that while income probably is a good index of socio-economic status at the extremes, income and socio-economic status are not so closely related in the middle ranges in which the bulk of the population lies.[2]

---

[1] Joseph A. Kahl and James A. Davis, "A Comparison of Indexes of Socio-economic Status," American Sociological Review, Vol. 20 (June, 1955), pp. 317-25.

[2] Ibid., p. 322.

Kahl has identified six aspects or "dimensions" of social stratification.[1] Each one can be used to stratify a given population, i.e., to assign its members to some hierarchical order. Kahl's "dimensions" are:

personal prestige
occupation and occupational prestige
possessions or life style
social interaction
class consciousness
value orientations

Measures of each of these factors have been used to place individuals in the status hierarchy.

In using social class concepts economists should take care not to use these various dimensions interchangeably. Although they are related closely, each is a separate concept and a separate set of techniques has been developed for classifying individuals within each dimension. The familiar six class division:

upper-upper
lower-upper
upper-middle
lower-middle
upper-lower
lower-lower

was developed by W. Lloyd Warner as a result of his studies which focussed chiefly on personal prestige and social interaction.[2] Some indexes developed for other dimensions have avoided the use of class or status categories and have merely assigned index scores. An example of this approach is the Index of Socio-Economic Status based on occupation, which is purported to be an index of occupational prestige rather

---

[1] Joseph A. Kahl, The American Class Structure (New York: Rinehart, 1957), pp. 8-12.

[2] W. Lloyd Warner, Marchia Meeker and Kenneth Eells, Social Class in America: The Evaluation of Status (New York: Harper, 1960), pp. 11-21.

than a measure of social class.[1]

The particular aspect or dimension which would seem to be of greatest interest for the study of consumer behavior is life style. It seems likely that the set of possessions of a family and its style of living would be an important determinant of its allocation of income. The extent of differences in expenditure attitudes and life styles between the classes probably has not been appreciated sufficiently by economists. The work of Pierre Martineau discussed in his article, "Social Classes and Spending Behavior," and the work of Lee Rainwater, Richard P. Coleman, and Gerald Handel discussed in their book, Workingman's Wife, suggest fundamental differences in life style, family goals and expenditure attitudes between the working and middle class.[2]

Although life style is the dimension which would be of most interest in the study of the effects of social status on consumer behavior, the indexes designed to stratify families by life style have proven less satisfactory than other measures of social status.[3] Scales based on the scoring of the contents of the family living room were much used in the 1930's as a device to measure social status. The rapid change in consumer tastes, the standardization of tastes and other problems have led to the abandonment of such measures.

---

[1] Albert J. Reiss, Jr., Otis Dudley Duncan, Paul K. Hatt and Cecil C. North, Occupations and Social Status, (New York: Free Press, 1961), pp. 139-51.

[2] Pierre Martineau, "Social Classes and Spending Behavior," Journal of Marketing, Vol. 23 (October, 1958), pp. 121-30.
Lee Rainwater, Richard P. Coleman, and Gerald Handel, Workingman's Wife: Her Personality, World and Life Style (New York: Oceana, 1959).

[3] Leonard Reissman, Class in American Society (Glencoe, Illinois: Free Press, 1959), pp. 117-25.

Since no suitable scales for the measurement of life style were available it was necessary to use another scale as a proxy variable. The scale chosen was the Index of Socio-Economic Status, which is based on the prestige scores of individual occupational categories. This substitution, while somewhat unsatisfactory from a conceptual viewpoint, was regarded as acceptable because of the identification, by Kahl and Davis, of occupation as an underlying factor accounting for much of the intercorrelation between different socio-economic status indexes.[1]

Thus, in this study a measure of occupational prestige will be used to determine social status. It is expected that the resultant status order will be much the same as that which would have been obtained had a life style measure of social status been employed. The development of the social class variable is discussed in Appendix III.

There is some reason to expect differences in working class and middle class food expenditure behavior. In their study, Workingman's Wife, Rainwater, Coleman and Handel found the working class wife to be concerned chiefly with cooking to please her family and found her to have little interest in either nutrition or the elaborate dishes suggested by the culinary experts of the women's page. This contrasts with the greater interest of middle class women in nutrition and in new recipes.[2] It was hypothesized that these differences in interests would be reflected in differences in food expenditure at the same level of income.

Index scores of Socio-Economic Status were assigned on the basis of the household head's occupation. On the basis of these scores households

---

[1] Kahl and Davis, p. 320-321.

[2] Rainwater, Coleman and Handel, pp. 178-84.

were assigned to one of four social status categories which could be described in the terminology of Warner's class system as: (1) upper-upper, lower-upper and upper-middle class (2) lower-middle class (3) upper-lower class (4) lower-lower class. Households whose heads were retired, unemployed, or not in the labor force were assigned to three separate categories since their Socio-Economic Status Index scores could not be determined. These three groups were: (5) unemployed (6) retired (7) not in the labor force (this group included housewives, students, etc.). The details of this process are discussed in Appendix III.

The omitted category in the analysis was the lower-middle class category.

The social class characteristic was found to be significant at the five percent level. A difference of $2.53 was found to exist between the geometric mean expenditure of all households and those of households whose head was not in the labor force taken at the geometric means of income and SMU. As a result of these tests the social class characteristic was judged to be significant in both a statistical and an economic sense.

Education of the Homemaker

Several different hypothesized causal relationships have underlaid the use of education of homemaker as a variable in the study of household food expenditure. The variable has been used both to study the effect of better nutritional knowledge on expenditure patterns and also as a proxy variable for social class.

Little evidence is available on the effect of education on food expenditures. In his 1950 study of food expenditures of Lansing

households, Thomas N. Moss found that education of housewife did not affect per capita food expenditure significantly when family size and income were controlled.[1] In his study of household expenditures for meat, Charles Zwick did not find significant differences in income elasticities between educational levels.[2] Zwick also found that price elasticities did not appear to be influenced by educational level. On the basis of these findings it was hypothesized that education of homemaker would not prove to be a significant variable.

Six educational level categories were employed:

1. No formal education or elementary school begun but not completed
2. Elementary school (8th grade) completed, no furtner education
3. High school begun but not completed
4. High school completed, no further education
5. College begun, but not completed
6. College, 4 or more years

The category omitted was "high school completed, but no further education."

The education of homemaker characteristic was found to be significant at the five percent level. The category whose geometric mean expenditure differed the most from the overall geometric mean expenditure was "no formal education or elementary school begun but not completed." The estimated mean expenditure in this group was $3.05 less than the overall mean for all households taken at the geometric means of income and SMU.

These results differ from the findings of both the Moss and Zwick studies. It was suspected that the results for this study might have been affected by correlated biases. The Moss and Zwick studies were both

---

[1] Moss, pp. 140-41.

[2] Zwick, pp. 452-55.

of urban households in a single region; it was believed that the results in this study might have been affected by differences in the distribution of regional and urbanization characteristics between educational level categories. The effects of these correlated biases will be considered in the next section of this chapter.

Self-employment of the Head

Studies of the economic behavior of households whose heads are self-employed have found substantial differences in savings behavior between these households and those whose head is not self-employed. Klein has found the marginal and the average propensities to save of unincorporated business owners and farmers to be markedly higher than those of all other households.[1] These differences in saving behavior can be interpreted alternatively as differences in consumption expenditure behavior, with the self-employed showing low marginal propensities to consume in comparison to other households.

Differences in food expenditure behavior between the households of the self-employed and other households have been noted in one study. Jean Crockett, in her study of data from the 1950 BLS Survey found the income elasticities for food of households with self-employed heads to be lower than those of other households.[2]

In his consideration of entrepreneurial saving, Klein preferred the classification of households into four categories for study:[3]

---

[1] Lawrence R. Klein, "Entrepreneurial Saving", Proceedings of the Conference on Consumption and Saving, Vol. II, pp. 317-19.

[2] Crockett, pp. 309-10.

[3] Klein, p. 299.

(1) self-employed businessmen (unincorporated only) (2) other self-employed (professional and skilled craftsmen) (3) employees (4) retired and unemployed persons. He felt that the behavior of the first two self-employed groups would differ in that the first group were true entrepreneurs whose businesses have larger capital requirements than members of the second group. He also felt that a distinction should be made between employees and the retired or unemployed, as it might be expected that retired and unemployed persons would be dissaving. The combination of the retired and unemployed with the regularly employed would confound the comparison of employees and the self-employed.

In this study, only two categories were used, although a scheme like the one employed by Klein would have been more desirable. Self-employment status was not ascertained in the original survey. It was possible, however, to determine self-employment status on the basis of the occupation reported for the household head. Although the classification obtained by this method is not so accurate as would have been obtained with a direct question, it was felt that it was, nevertheless, reasonably accurate.

The two categories employed are self-employed and not self-employed. Not self-employed was the omitted category.

The characteristic was significant at the five percent level. The geometric mean of expenditures of households whose head was self-employed was $3.28 less than that for households whose head was not self-employed at the geometric mean of income and SMU. The characteristic was judged significant in both a statistical and an economic sense.

It was felt that correlated biases arising from the presence of a large number of farmers in the self-employed group might have affected the results obtained. The presence of such biases will be investigated later in this chapter.

Employment of the Homemaker

It is supposed generally that households whose homemaker is employed spend more for food than similar households in which the housewife is not employed. This difference is presumed to exist because it is expected that the employed homemaker will utilize more expensive time-saving processed foods than the homemaker who is not employed.

There is some evidence that expenditures for food are higher in households in which the homemaker is employed. Cross-classifications of this data from the 1955 U.S.D.A. Survey indicate that average weekly expenditures per person for food consumed at home were $7.66 for households whose homemaker was employed and $6.49 in households whose homemaker was not employed.[1] It must be emphasized, however, that the households of homemakers who were not employed included more children under 16 giving rise to the possibility of lower costs per person. These households also were, on the average, larger than those in which the homemaker was employed, giving rise to the possibility of economies of scale.

All homemakers who were employed outside the home, either part-time or full-time, were classified in the "Employed Outside the Home" category. The omitted category was "Not Employed Outside the Home." The employment characteristic was not significant at the five percent level. The difference between the geometric mean expenditure of all households and households in which the homemaker was employed was $.58, taken at the geometric means of income and SMU. The characteristic was judged to be non-significant in both a statistical and an economic sense. It is

---

[1] Janet Murray, "Food Consumption and Dietary Levels of Households with Employed and Non-employed Homemakers," U.S. Department of Agriculture, Family Economics Review (June, 1960), pp. 9-13.

believed that most of the differences in expenditures noted in earlier studies between households of employed homemakers and those of not employed homemakers were due to income and compositional differences between the two sets of households. With these differences taken into account, the differences in expenditures do not appear to be of significance.

Determination of the Effects of Correlated Biases

As has been noted in the previous section, it was suspected that correlated biases might have affected the results of several of the tests of significance. The absence of controls for the effects of urbanization and region could have been a major source of such biases.

It was noted earlier that the strategy of testing each characteristic separately was chosen to avoid the adverse effects of multicollinearity on the tests of significance. Some technique is necessary, however, to deal with the effects of correlated biases if characteristics are tested separately, one at a time. In order to deal with suspected sources of correlated biases certain further tests of significance were undertaken. In these tests certain suspected sources of correlated biases are controlled so that the significance of particular variables can be evaluated free from the correlated biases produced by the controlled variables.

The Tests Employed

The tests employed are the same basically as those employed in the previous section. The characteristic under test and the characteristics suspected to be the source of correlated biases will be represented by dummy variables. The equation form including these variables will be:

$$\log Y = a + b \log X_1 + c \log X_2 + \sum_{i=2}^{I_1} d_{1i} Z_{1i} + \sum_{i=2}^{I_2} d_{2i} Z_{2i} + \sum_{i=2}^{I_3} d_{3i} Z_{3i} + U \qquad (2.12)$$

where Y, X_1 and X_2 are defined as before and the Z_{1i}, Z_{2i} and Z_{3i} are sets of dummy variables representing three sets of characteristics having I_1, I_2 and I_3 categories respectively. To test the significance of the characteristic Z_1, with characteristics Z_2 and Z_3 controlled, SSR_{Ho} is estimated for an equation of form (2.12) which omits the Z_{1i} variables. The SSR_{Ha} and SSE_{Ha} are estimated in an equation of form (2.12). These values are employed in the F test discussed previously.

The same test of economic significance will be employed. A difference in expenditure of $1.00 or more between one category and the overall mean expenditure with income and SMU held constant will be required for a characteristic to be judged significant in an economic sense.

Analysis and Results

Urbanization

Since urbanization was the most significant single variable in explaining the variance of expenditures it was expected that it would be a major source of correlated biases affecting the tests of other variables. The test for the significance of urbanization was not expected to be much affected by correlated biases. It did seem desirable to determine the possible effects of correlated biases arising from the absence of controls for regional effects, however.

The significance of urbanization was tested in a regression similar to (2.12), which included dummy variables representing three of the four regions. Urbanization was found to be significant at the five percent

level with the effects of region, income and SMU controlled. The geometric mean expenditure of rural farm households was $6.94 less than the mean expenditure of all households with the effects of region, income and SMU all controlled. The urbanization characteristic was judged to be of major significance, even after the effects of region were controlled.

The adjusted coefficient of multiple determination, $\bar{R}^2$ increased from .510 in the equation which included income, SMU and region variables to .564 in the one which also included the urbanization variables.

The details of the F tests for the further tests of significance which are discussed in this section are presented in Table 2.4.

Region

Region was considered to be the second most important variable in explaining variation of expenditures between households. It was desired to determine the significance of region after the key variable, urbanization, was controlled.

Regionality was found to be significant at the five percent level with the effects of urbanization, income and SMU controlled. The geometric mean expenditure of households in the Southern region was $2.04 less than the mean for all households after the effects of urbanization, income and SMU were controlled. Region was judged to be highly significant, even after control of urbanization effects.

The $\bar{R}^2$ obtained increased from .553 in a regression including income, SMU and urbanization variables to .564 with the addition of the regionality variables.

Race

It was desired to determine whether correlated biases arising from

Table 2.4 Tests for the Effects of Correlated Biases

Characteristic Under Test	n	S-D Variables Controlled	SSR_{Ha}	SSE_{Ha}	SSR_{Ho}	Test Statistic	$F_{.05}$	Computed F
Urbanization	3641	Region	126.83	97.86	114.67	$F_{.05}(?,3633)$	2.99	225.79**
Region	3641	Urbanization	126.83	97.86	124.33	$F_{.05}(3,3633)$	2.60	30.88**
Race	3596	Urbanization	122.37	98.35	121.31	$F_{.05}(1,3590)$	3.84	38.77**
Race	3596	Region	112.07	108.65	112.06	$F_{.05}(1,3589)$	3.84	.36
Life cycle	3641	Urbanization, region	128.89	95.81	126.83	$F_{.05}(10,3623)$	1.83	7.79**
Social class	3641	Urbanization	125.63	99.07	124.33	$F_{.05}(6,3630)$	2.09	7.93**
Education of the homemaker	3469	Urbanization, region	119.43	90.83	117.43	$F_{.05}(5,3456)$	2.21	15.21**
Self-employment of the head	3641	Urbanization	124.48	100.21	124.33	$F_{.05}(1,3635)$	3.84	5.45*

regional or urbanization effects had affected the test of significance for the race characteristic.

The race characteristic was first tested with urbanization controlled. It was found to be significant at the five percent level with urbanization, income and SMU controlled. The geometric mean expenditure of non-white households was found to be $2.67 less than the expenditure of all households with urbanization, income and SMU controlled. Race was judged to be significant in both a statistical and an economic sense with the effects of urbanization controlled.

The race characteristic was also tested with region controlled. Race was <u>not</u> found to be significant at the five percent level after regional, income and SMU effects were controlled. The geometric mean of expenditures of non-white households was found to be $.15 less than the mean expenditure of all households with the effects of region, income and SMU controlled.

The $\bar{R}^2$ obtained for a regression including income, SMU and urbanization increased from .549 to .554 with the addition of the race variable. The $\bar{R}^2$ obtained for the regression including income, SMU and region was not affected by the addition of the race variable.

Stage in the Family Life Cycle

It was desired to determine whether correlated biases arising from regional and urbanization effects had influenced the test of significance for stage in the family life cycle. The effects of stage in the family life cycle were evaluated while controlling for the effects of both region and urbanization.

Stage in the family life cycle was found to be significant at the five percent level, with the effects of urbanization, region, income and

SMU controlled. The geometric mean expenditure of the category with "male or female head over age 65" was $3.32 less than the mean for all households with the effects of region, urbanization, income and SMU controlled. The characteristic was judged still to be significant after regional and urbanization effects were controlled. The $\bar{R}^2$ obtained increased from .564 for the regression including the income, SMU, urbanization and region variables to .572 when the variables representing stage in the family life cycle were added.

Social Class

The social class characteristic was also examined to determine the possible effects of correlated biases arising from urbanization effects. There was reason to suspect that the assignment of all farmers to a single class category might have affected the results.

The social class characteristic was found to be significant at the five percent level, with the effects of urbanization, income and SMU controlled. The geometric mean of expenditure in the "not in the labor force" category was $2.78 less than the overall mean for all households after the effects of urbanization, income and SMU were taken into account. The characteristic was judged to be both statistically and economically significant with the effects of urbanization controlled.

The $\bar{R}^2$ obtained increased from .553 in a regression in which income, SMU and urbanization variables were included to .553 in one which also included social class variables.

Education of Homemaker

The effects of correlated biases arising from urbanization and region on the test of the significance of the education of the homemaker were examined. It was suspected that the lower levels of education

prevalent in the South and rural areas might have affected the results. The effect of education of the homemaker was tested while controlling for both urbanization and region at the same time.

Education of the homemaker was found to be significant at the five percent level, with region, urbanization, income and SMU controlled. The geometric mean expenditure of the "elementary school not completed or no formal education" category was $2.89 less than the mean expenditure of all households after the effects of urbanization, region, income and SMU were taken into account. The education characteristic was judged to be significant in both a statistical and an economic sense.

The $\bar{R}^2$ obtained increased from .558 for the regression including income, SMU, urbanization and region variables to .566 with the addition of the education of homemaker variables.

Self-Employment of the Head

The effects of correlated biases arising from the urbanization characteristic on the test of significance for self-employment were examined. It was expected that the results had been affected substantially by the inclusion of virtually all farm households in the self-employed category.

The self-employment characteristic was significant at the five percent level, with urbanization, income and SMU controlled. The geometric mean expenditure of households with self-employed heads was $1.05 more than the mean expenditure of all households after the effects of urbanization, income and SMU were taken into account. Controlling for the effects of urbanization has thus changed the estimate of the difference in the expenditures of self-employed households from the overall mean from -$3.28 to +$1.05.

The $\bar{R}^2$ obtained increased only from .5528 for the regression in which income, SMU and urbanization were included to .5534 with the addition of the self-employment variable.

The changes in the estimate of the effect of self-employment produced by controlling for urbanization provide a further illustration of the impact of correlated biases. In the original test households with self-employed heads were estimated to spend $3.28 less than the overall mean expenditure for all households. After the effects of urbanization were taken into account the households with self-employed heads were estimated to spend $1.05 more than the mean expenditure of all households. This dramatic shift shows the impact of correlated biases on both tests of significance and parameter estimates. It also gives rise to some concern about the reliability of the technique used to determine self-employment status.

Self-employment status was not ascertained in the survey schedule of the 1955 USDA study. In order to determine self-employment status for the present study the author utilized the reported occupation of the household head. Households whose head's occupation was one in which workers are usually self-employed were coded as self-employed. Most farmers and professionals were coded as self-employed unless there were clear indications to the contrary. Those reporting other occupations which are not typically ones in which workers are self-employed were coded as not self-employed unless there was clear indication of self-employment. The results suggest that the preponderance of those assigned to the self-employed category were farmers and professionals. There is reason to believe that many households may not have been assigned correctly. Occupation reported was the only basis on which to base assignment to the self-employed or not self-employed categories.

It does not seem to provide sufficient information for accurate coding. Clearly, a direct question about self-employment status is necessary for accurate assignment. On the basis of the results obtained it seems unwise to make a final judgment on the significance of the self-employment characteristic. The problem merits further study when suitable survey data become available.

Conclusions About the Specification of the Structural Model

Having tested the statistical and economic significance of the eight S-D variables, we are now in a position to specify a structural model explaining household food expenditures. Let us first review the results of the tests which have been performed, however.

The eight S-D variables were examined in an initial round of tests in which the significance of each S-D variable was tested separately with the effects of income and SMU held constant. The employment of the homemaker variable was judged non-significant. All seven other variables were judged to be significant in both a statistical and an economic sense.

In the second round of tests, which were aimed at the elimination of the effects of correlated biases, race was judged to be non-significant when region is taken into account. No final judgment was made about the significance of self-employment status because of the unexpected results obtained when the significance of self-employment was tested with urbanization controlled.

On the basis of the tests and the problems dealt with in developing the preliminary economic model we can specify the structural model in a more detailed form:

Food expenditure = f (Income, SMU, Urbanization, Region, Life cycle stage, Education of the homemaker, Social class).

In the next chapter we shall proceed to consider the problems of developing measures of the five S-D variables included in the structural model.

Before proceeding, however, it seems worthwhile to attempt to rank the five S-D variables included in the structural model in the order of their importance in explaining variance in expenditure between households. As could be seen in the previous section some variables add relatively little to the explanation of the variance of expenditure provided by the variables income, SMU, urbanization and region. It can be seen in Table 2.5, for example, that once regional effects are taken into account, the racial characteristics variable adds nothing to the explanation. The addition of the race variable produces no change in the estimate of $\bar{R}^2$.

From the $\bar{R}^2$ values presented in Table 2.5 we can estimate the approximate contribution of each of the five significant variables to an explanation of the variance of expenditure. Using these estimates of contribution to $\bar{R}^2$ as the criterion of relative importance, the variables rank as follows:

urbanization	.054
region	.011
life cycle stage	.008
education of homemaker	.008
social class	.005

As can be seen, the contributions of all the variables are rather small relative to the 49.5 percent of the variance in expenditure explained by income and SMU. If the 5 variables above are completely

Table 2.5 Estimated Adjusted Coefficients of Multiple Determination ($\bar{R}^2$) [a]

S-D Variable	$\bar{R}^2$ under null hypothesis. Income, SMU and variable below included in regression.			$\bar{R}^2$ under alternate hypothesis. Income, SMU, variable below and S-D variable included in regression.		
	Urbanization	Region	Urbanization & region	Urbanization	Region	Urbanization and region
Urbanization	--	.510	--	--	.564	--
Region	.553	--	--	.564	--	--
Race	.549	.507	--	.554	.507	--
Stage in family life cycle	--	--	.564	--	--	.572
Social class	.553	--	--	.558	--	--
Education of homemaker	--	--	.558	--	--	.566
Self-employment of head	.5528	--	--	.5534	--	--

[a] Instances in which information was not computed or is not applicable are denoted by "--".

independent of each other, they may add as much as 8.6 percent to the explained portion of the variance of food expenditure. Because of known intercorrelations, the use of the 5 variables, would probably increase the $\bar{R}^2$ by about .075 above the $\bar{R}^2$ obtained for a regression including only income and SMU variables. Despite the small size of their individual contributions to $\bar{R}^2$, it is felt that as a group their total contribution is of some importance.

CHAPTER III

Specification of Socio-economic and Demographic Variables

The regression results, in addition to providing a basis for the tests of significance, permit us to examine the effects of the individual S-D variables in some detail. Given the regression results we can (1) determine whether the effects of a particular S-D characteristic fall into some pattern and (2) determine which categories of the characteristic differ significantly in their expenditure behavior. On the basis of such information we should be able to refine the specification of the significant S-D variables. Because of their essentially qualitative nature, the S-D characteristics do not appear likely to lend themselves to representation by continuous variables. However, since the flexibility of continuous variables makes their use desirable wherever feasible, the possibility of representing S-D characteristics by continuous variables does bear investigation.

Examination of the Effects of the Variables

Analytical Procedure

In examining the coefficients estimated for the individual S-D categories we will first determine the significance of the differences in expenditure between categories and then will consider the overall pattern of effect of the categories making up a S-D characteristic.

In order to interpret the coefficients of the S-D category variables correctly we must remind ourselves that they represent the difference in expenditure between the omitted category and the S-D category under consideration. In a model of the effects of a single S-D characteristic in which all T categories are included, such as

$$Y_t = \sum_{i=1}^{I} a_i Z_{it} + u_t \qquad t=1,\ldots n \qquad (3.1)$$

where

$Z_{it} = 1$ when observation t falls in category i

$= 0$ otherwise

the a_i are the effect on expenditure of membership in category i. In contrast, in the model in which one category is omitted and a constant term added

$$Y_t = a_o + \sum_{i=2}^{I} d_i Z_{it} + u_t \qquad (3.2)$$

where Z_{it} is defined as above, a_o is, in effect, an estimate of the coefficient of the omitted category, i.e., an estimate of a_1.

In model (3.1) for observations in category i

$$Y_{it} = a_i + u_{it} \qquad i = 1,\ldots I \qquad (3.3)$$

and in model (3.2) we have for observations in category i

$$Y_{it} = a_o + d_i + u_{it} \qquad i = 2,\ldots I \qquad (3.4)$$

We can see that a_i must equal $a_o + d_i$ and that

$$a_i - a_o = d_i$$

that is, d_i is an estimate of the difference in effect between the omitted category and category i.[1]

---

[1] Robert L. Gustafson, "The Use and Interpretation of 'Dummy Variables' in Regressions," Note of January 22, 1962 (Revised), Agricultural Economics Department, Michigan State University. (Mimeo.)

A test of the hypothesis $d_i = 0$ about the coefficient of any individual category is, then, equivalent to a test of the hypothesis $a_i - a_1 = 0$, where a_1 is the effect of the omitted category. In this way we can test whether the effect of membership in category i is significantly different from that of membership in the omitted category.

The test statistic for the test of the hypothesis $d_i = 0$ is

$$\frac{\hat{d}_i}{S_{di}}$$

It has a t distribution with (n-k-1) degrees of freedom, where k is the number of regression coefficients estimated, where $\hat{d}_i$ is the estimate of d_i and S_{di} is the standard error of the regression coefficient $\hat{d}_i$.[1]

In addition to testing whether the effect of membership in a given category is significantly different from membership in the omitted category we also can test whether the effects of membership in any two given categories included in the regression analysis are significantly different. This test is equivalent to a test of the hypothesis

$$a_i - a_j = 0 \qquad \begin{array}{l} i = 2,\ldots I \\ j = 2,\ldots I \\ i \neq j \end{array}$$

Now we have seen that

$$a_i = d_i + a_o$$

$$a_j = d_j + a_o$$

---

[1] George W. Snedecor, Statistical Methods (5th ed.; Ames, Iowa: Iowa State College Press, 1959), pp. 418-19.

thus, a test of the hypothesis

$$a_i - a_j = 0$$

is equivalent to a test of the hypothesis

$$(d_i + a_o) - (d_j + a_o) = 0$$

or

$$d_i - d_j = 0$$

This latter hypothesis can be tested by the usual t test employed for testing the significance of differences between regression coefficients in the same equation. The test statistic is

$$\frac{\hat{d}_i - \hat{d}_j}{\sqrt{S^2_{y.123\ldots k}\,(C_{ii} + C_{jj} - 2C_{ij})}}$$

it has a t distribution and has n-k-1 degrees of freedom, where k is the number of regression coefficients estimated, $S^2_{y.123\ldots k}$ is the squared standard error of estimate and C_{ii}, etc., are elements of the inverse matrix.[1]

Results for the Individual S-D Variables

Urbanization

The coefficients estimated for the urbanization categories in an equation also including regional categories are given in Table 3-1.

---

[1] The test of differences between regression coefficients in the same equation is discussed in Snedecor, p. 442.

Table 3-1. Regression Results for Urbanization and Region

Variable [a]	Estimated Coefficient [b]	Computed t	Percent Expenditures Are of Those of Omitted Category [c]
Constant term	-.75208 (.04064)	-18.508**	--
Income	+.29269 (.01124)	+26.044**	--
SMU	+.61138 (.01272)	+48.125**	--
Rural non-farm	-.05034 (.00655)	- 7.778**	89.0
Rural farm	-.18319 (.00390)	-21.141**	64.8
Northeast	+.01967 (.00711)	+ 2.765**	104.7
South	-.04765 (.00724)	- 6.577**	89.7
West	+.01439 (.00945)	+ 1.523	103.4

n = 3641

$\bar{R}^2$ = .564

$\bar{Y}$ = $22.38

S_y^2 = .02694

$\overline{\log Y}$ = 1.29551

a See Chapter 2 for detailed definitions.

b The Standard errors of the estimated regression coefficients are given in parentheses.

c Instances in which this information was not applicable or not computed are denoted by "--".

* Significant at a probability level between 5 percent and 1 percent.

** Significant at the 1 percent level or less.

The difference in level of expenditure between the rural non-farm category and the omitted urban-North Central category was found to be significant at the one percent level. The difference between rural farm expenditures and urban-North Central expenditures was also found to be significant at the one percent level. The difference between the two rural categories also was found to be significant at the one percent level. The computed t value for the test of this difference was 14.46.

The effect on the coefficients estimated of intercorrelations with the regional variables can be seen in Table 3.2. The unadjusted coefficients are those estimated in a regression which did not include regional effects, the adjusted coefficients (adjusted in this case for the effects of region) are those estimated in the regression whose results are presented in full in Table 3.1. The adjustment for region decreased the negative effect of both the rural non-farm and the rural farm categories. This change apparently was due to the removal of biases arising from the intercorrelation of Southern regional residence and membership in the rural categories.

Region

The coefficients estimated for the regional categories are given in Table 3.1. The coefficients were estimated in an equation which also included urbanization categories.

The difference in level of expenditures between the Northeast category and the omitted urban-North Central was significant at the one percent level. The difference between the South and the urban-North Central categories was also significant at the one percent level. The difference between the West and urban-North Central categories was not significant at the five percent level. The difference is significant

TABLE 3-2. Effects of Intercorrelations with Region on the Coefficients Estimating Urbanization Effects

Urbanization Category	Unadjusted Coefficient[a]	Adjusted Coefficient[b]
Rural non-farm	-.05495	-.05094
Rural farm	-.19278	-.18819

a Estimated in regression not including regional variables.

b Estimated in regression including regional variables.

at the ten percent level, however. The difference between the coefficients of the Northeast and the West was not significant at the five percent level. The computed t for this test was .56.

The results indicate that the division of the country into South and non-South categories which has been employed in some studies is a suitable rough approximation for dealing with the effects of region. The results do suggest that it would be preferable to retain the four categories employed in the 1955 USDA Survey, however.

The effects of biases arising from intercorrelations of region with urbanization on the coefficients estimated can be seen in Table 3.3. It appears that the high proportion of urban residents in the Northeast and the West biased the unadjusted coefficient upwards.

Race

The coefficient estimated for the effect of membership in the non-white category is given in Table 3.4. The coefficient was estimated in a regression which also included regional variables. The 3596 households reporting race were included in this analysis. Some 320 of these were non-white households, principally negro. It will be recalled that with regional effects taken into account the race characteristic was found to be not significant in either a statistical sense, at the five percent level, or in an economic sense.

The expenditures of non-white households were estimated to be 99.3 percent of those of the mean expenditure of all households. This difference in expenditures was not found to be significant at the five percent level.

The effects of the correlated biases arising from the effects of region and urbanization were pointed out earlier. The extent to which they affected the estimates of effect of membership in the non-white

TABLE 3-3. Effects of Intercorrelations with Urbanization on the Coefficients Estimating Regional Effects

Regional Category	Unadjusted Coefficient[a]	Adjusted Coefficient[b]
Northeast	.04099	.01967
South	-.03816	-.04765
West	.03405	.01439

a Estimated in regression not including urbanization variables.

b Estimated in regression including urbanization variables.

TABLE 3-4. Regression Results for Race

Variable	Estimated Coefficient	Computed t	Percent Expenditures Are of Those of Omitted Category
Constant term	-.89186 (.04305)	-20.718**	--
Income	+.34854 (.01175)	29.664**	--
SMU	+.54814 (.01318)	41.586**	--
Non-white	-.00316 (.01099)	- .288	99.3
Northeast	+.04294 (.00740)	5.802**	--
South	-.03728 (.00804)	- 4.639**	--
West	+.03555 (.00993)	3.580**	--

$n = 3596$

$\bar{R}^2 = .51$

$\bar{Y} = \$22.91$

$S_y^2 = .03027$

$\overline{\log Y} = 1.29646$

category can be seen in Table 3.5. Adjustment for urbanization substantially increased the negative effect of membership in the non-white category, while adjustment for region reduced the coefficient for the category to the point where it was no longer significant.

Stage in the Family Life Cycle

The life cycle stage categories employed and the number of households in each category are presented in Table 3.6.

The coefficients estimated for the stage in the family life cycle categories are given in Table 3.7. These estimates were obtained in a regression which also included variables for urbanization and region.

Of the 10 categories employed only 5 were found to have expenditure levels which were significantly different from those of the omitted category, "Couple, oldest child 6-15." Those found to be significantly different were chiefly households with either a male or female head without children and older households. The results have been arrayed in Figure 3.1 in order to determine whether any patterns of effects can be observed.

The pattern of the array by life cycle stage suggests that there is little difference in expenditure pattern between the various types of husband and wife categories in which children are present. The expenditures of the two younger husband and wife household categories without children (Categories 1 and 6) are also very similar to this pattern. The category of households with either a male or female head, under age 65, with children present (category 9) also behaves in a similar fashion to the husband and wife households with children.

The two categories of households with either a male or female head in which no children were present (categories 7 and 8) had very similar estimated coefficients, which were significantly different from the

TABLE 3-5. Effects of Intercorrelations with Urbanization and Region on the Coefficient Estimating Race Effects

Racial Category	Unadjusted Coefficient[a]	Coefficient with Adjustments for Urbanization[b]	Coefficient with Adjustments for Region[c]
Non-white	-.02797	-.06292	-.00316

a Estimated in a regression including variables for income, SMU and non-white category membership.

b Estimated in a regression including income, SMU, urbanization and race as independent variables.

c Estimated in a regression including income, SMU, region and race as independent variables.

TABLE 3-6. Descriptions of the Life Cycle Categories with Number of Observations in Each

Life Cycle Category Number	Category Description	n
1	Couple, wife under 40 no children	200
2	Couple, oldest child under 6	487
(3)[a]	Couple, oldest child 6-15	940
4	Couple, oldest child 16-19	651
5	Couple, wife 40-64 young adults 20-39	140
6	Couple, wife 40-64 no children present	527
7	Male or Female Head, under 50 no children present	91
8	Male or Female Head, 50-64 no children present	144
9	Male or Female Head, under 65 children present	135
10	Couple, wife 65 or over no children present	171
11	Male or Female Head, age 65 or over children may be present	155

a This category was the one omitted in the regression analysis.

TABLE 3-7. Regression Results for Stage in the Family Life Cycle

Life Cycle Category Number	Variable[a]	Estimated Coefficient	Computed t	Percent Expenditures Are of Those of Omitted Category
	Constant term	-.59377 (.05399)	-10.998**	--
	Income	+.27489 (.01177)	+23.351**	--
	SMU	+.56718 (.01910)	+29.699**	--
1	Couple, wife under 40 no children present	-.01353 (.01384)	- .977	96.9
2	Couple, oldest child under 6	+.00935 (.00936)	.999	102.2
4	Couple, oldest child 16-19	-.02326 (.00841)	- 2.766**	94.8
5	Couple, wife 40-64 young adults 20-39	-.01348 (.01497)	- .901	96.9
6	Couple, wife 40-64 no children present	-.01688 (.01046)	- 1.614	96.2
7	M or F Head, under 50 no children present	-.04074 (.02015)	- 2.021*	91.0
8	M or F Head, 50-64 no children present	-.04917 (.01776)	- 2.769**	89.3
9	M or F Head, under 65 children present	-.00036 (.01539)	- .024	99.9
10	Couple, wife 65 or over No children present	-.08617 (.01472)	- 5.853**	82.0
11	M or F Head, age 65 or over children may be present	-.09352 (.01671)	- 5.596**	80.6
	Rural non-farm	-.05112 (.00651)	- 7.851**	--
	Rural farm	-.18525 (.00892)	-20.773**	--
	Northeast	+.02031 (.00705)	+ 2.880**	--
	South	-.05117 (.00722)	- 7.091**	--
	West	+.01381 (.00938)	+ 1.473	--

n = 3641

$\bar{R}^2$ = .57

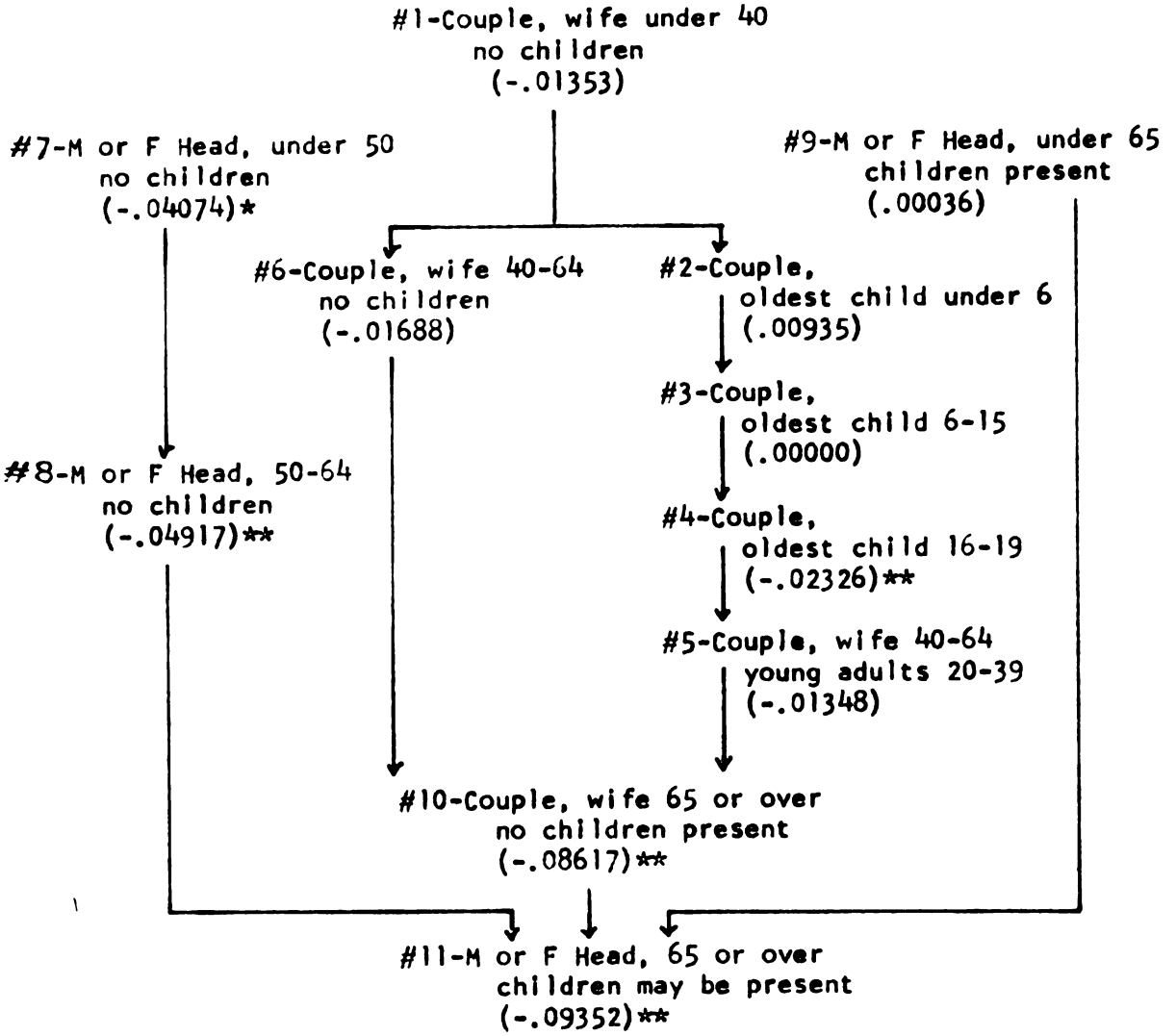
$\bar{Y}$ = $22.88

Sy^2 = .02644

$\overline{\log Y}$ = 1.29551

a See Chapter 2 for detailed definitions of variables.

FIGURE 3-1. Stages in the Family Life Cycle and Estimates of Their Effects on Food Expenditure



omitted category at the five percent level or better. These coefficients were found to be not significantly different from each other at the five percent level. The computed t for this test was .17.

The two categories of household with head age 65 or over (category 11) and wife age 65 or over (category 10) had the lowest coefficients of any of the life cycle categories. Both coefficients were significantly different from the omitted category at the one percent level. The two coefficients were found to be not significantly different from each other at the five percent level. The computed t for this test was .21.

The value of the SMU variable for each household takes into account both the total number of meals served to members of the household and the age-sex composition of the household. The values used for the age-sex scale are based on nutritional requirements rather than observed food consumption behavior. The SMU variable thus does take differences in requirements into account. Any differences between life cycle categories which remain after this adjustment must be due to stage in life cycle itself or to age-related differences which cannot be explained on the basis of differences in requirements.

The differences in expenditure for the two categories representing final stages in the family life cycle (categories 10 and 11) are unexpectedly large. The coefficients appear to be attributable chiefly to old age itself, since the most likely sources of correlated biases have been controlled by inserting the income, SMU, region and urbanization variables in the analysis.

The results viewed in their entirety suggest that it is important to consider all three of the dimensions employed in constructing the life cycle categories: marital status of head, age of head or wife, presence of children. The younger households headed by couples all

behaved rather similarly, whether or not children were present. The behavior of younger households with either a male or female head differed markedly depending on whether or not children were present. The households in which children were present behaved like the households headed by couples, while those in which no children were present spent significantly less than the households headed by couples. In this case, a combination of marital status, age, and the presence of children are needed to explain the differences. The third major life cycle stage group, the two final life cycle stages, differ from the other categories principally in the age dimension. Thus, all three dimensions employed come into play in explaining differences between the three major groups of life cycle categories.

Social Class

The seven social class categories employed are given in Table 3.8, along with the number of households classified in each category. The regression results are presented in Table 3.9. The coefficients presented were estimated in a regression which also included variables for urbanization.

On examining the coefficients for the upper and upper-middle class category, the upper-lower class category and the lower-lower class category, we find that only the coefficient of the lower-lower class category is significantly different from zero. The expenditure behavior of the upper and upper-middle class category and the upper-lower class category thus do not appear to differ significantly from that of the omitted lower-middle class category, once income and SMU are taken into account.

The coefficients of the three categories of households (Retired, Unemployed, Not in the labor force) which could not be given social

TABLE 3.3. Descriptions of the Social Class Categories Employed with Number of Observations in Each

	Variable	n
1	Upper and Upper-middle class	402
(2)[a]	Lower-middle class	908
3	Upper-lower class	1523
4	Lower-lower class	337
5	Retired	215
6	Unemployed	112
7	Not in the labor force	144
		3641

a Category omitted in the regression analysis.

TABLE 3.9. Regression Results for Social Class

Variable	Estimated Coefficient	Computed t	Percent Expenditures Are of Those of Omitted Category
Constant term	-.66859 (.04605)	-14.518**	--
Income	+.28096 (.01241)	+22.635**	--
SMU	+.59387 (.01304)	+45.542**	--
Upper and Upper-middle class	+.00621 (.01002)	+ .620	101.4
Upper-lower class	-.00939 (.00737)	- 1.274	97.8
Lower-lower class	-.04393 (.01091)	- 4.027**	90.4
Retired	-.05648 (.01318)	- 4.287**	87.8
Unemployed	-.05852 (.01698)	- 3.447**	87.4
Not in the labor force	-.06592 (.01556)	- 4.236**	85.9
Rural non-farm	-.05332 (.00659)	- 8.096**	--
Rural farm	-.19948 (.00926)	-21.536**	--

$n = 3641$

$\bar{R}^2 = .56$

$\bar{Y} = \$22.88$

$S_y^2 = .02729$

$\overline{\log Y} = 1.29551$

class assignments because the head did not report a regular occupation were all significant at the five percent level. The behavior of these three categories, therefore, can be considered to differ significantly from that of the lower-middle class, even after income and SMU are taken into account.

It appears that social class, as it is usually conceived, has little influence on total food expenditure after the effects of income have been controlled. It does appear, however, that whether or not one is employed regularly is of significance. The three categories of households whose heads were not employed regularly due to retirement, unemployment or not being in the regular labor force, all displayed similar behavior. The behavior of the lower-lower class households whose heads are in occupations where employment typically is irregular showed an interesting similarity to that of the three categories which did not have any regular occupations. The coefficient of the lower-lower class category was not significantly different from that of the unemployed category at the five percent level. The computed t for this test was .25.

The strongly negative coefficients of the three not regularly employed categories and the lower-lower class category perhaps may be explained in part by the prevalence of gifts of food and welfare-donated food for households of these types. It was suspected that a disproportionately large number of households in later life cycle stages were included in these four "deviant" categories. The unemployed and not in the labor force categories were suspected to contain a large number of older households; the presence of a large number of older households in the retired category is obvious. Because of these intercorrelations with stage in the family life cycle, the coefficients of the four "deviant"

categories may overstate the actual effects of membership in these categories.

The usefulness of the social class variable in economic analyses has often been questioned because of its purportedly high correlation with income. The highest correlation between social class category membership and log income found in this study was only .30. This was the simple correlation between income and membership in the upper and upper-middle class category. This rather low correlation supports the view that income and social class are not interchangeable concepts. The rather low correlation coefficient found also supports the view that multi-collinearity did not affect the analysis to any important extent.

On the whole, regular social class concepts do not appear to be particularly useful in explaining the variance of total food expenditures despite the fact that the concept cannot be regarded as interchangeable with income. Instead, the fact of regular employment or the head's membership in a regular occupation appears to be the aspect of social class which is of greatest importance. It should perhaps be noted that while regular social class concepts have not proven particularly satisfactory in explaining variation in total food expenditures it is likely that they would be somewhat more useful in explaining differences in expenditure on individual items or sub-categories of food.

Education of the Homemaker

The education of the homemaker categories employed and the number of observations in each category are presented in Table 3.10. Of the 3641 households included in this study 172 had no wife or female head present or did not report her education. These households were not included in this analysis.

TABLE 3.10. Descriptions of the Categories of Education of the Homemaker Employed with Number of Observations in Each

	Variable	n
1	Elementary school not completed or no formal education	455
2	Elementary school completed, no further education	748
3	High school not completed	685
(4)[a]	High school completed, no further education	1094
5	College not completed	294
6	College, 4 or more years	193
		3469

a Category omitted in the regression analysis.

The results of the regression analysis are presented in Table 3.11. The coefficients were estimated in a regression which also included variables for region and urbanization.

The categories with less education than the omitted category ("high school completed, no further education") had negative coefficients while the coefficients of the two categories with more education than the omitted category were positive. Only the coefficients of the two categories with less education that the omitted category were significant, however, indicating that their behavior was significantly different from that of households in which the homemaker had completed high school, but had no further education.

There is reason to believe that the effects observed for the two lowest educational categories, "elementary school not completed, or no formal education," and "elementary school completed, no further education" were due in part to a concentration of older households in these lower education categories. We know that in the adult population as a whole educational attainment is negatively correlated with age, membership in the negro race, and rural and Southern residence.[1] The use of urbanization and regional variables can be expected to deal with two of these potential sources of correlated biases. The effects of race on food expenditure have been judged non-significant. The effects of age have not been controlled, however. From our examination of the effects of life cycle stage we would expect a concentration of older households in the lower education categories to bias their coefficients downward. The possible effects of this source of bias should be determined.

---

[1] See for example, U. S. Bureau of the Census, "Educational Attainment: March 1962, "Current Population Reports, Population Characteristics, Series P-20, No. 121 (February 7, 1963).

TABLE 3.11. Regression Results for Education of the Homemaker

Variable	Estimated Coefficient	Computed t	Percent Expenditures are of Those of Omitted Category
Constant term	-.62943 (.04466)	-14.093**	--
Income	+.25705 (.01244)	+20.668**	--
SMU	+.62168 (.01323)	+46.997**	--
Elem. school not completed or no formal education	-.06782 (.00971)	- 6.986**	85.5
Elem. school completed, no further education	-.02987 (.00788)	- 3.788**	93.4
High school not completed	+.00146 (.00802)	+ .182	100.4
College not completed	+.01567 (.01073)	+ 1.460	103.7
College, 4 or more years	+.02030 (.01282)	+ 1.583	104.8
Rural non-farm	-.05271 (.00663)	- 7.952**	--
Rural farm	-.19224 (.00908)	-21.162**	--
Northeast	+.02267 (.00721)	+ 3.146**	--
South	-.04217 (.00746)	- 5.655**	--
West	+.00817 (.00971)	+ .841	--

n = 3469
$\bar{R}^2$ = .57
$\bar{Y}$ = $23.07

Sy^2 = .02628
$\overline{\log Y}$ = 1.30043

There was reason to believe that social class and education of the homemaker might be, at least to some extent, interchangeable variables in explaining the variance of food expenditures. The patterns of results for both variables show them to have significant effects for apparently rather similar groups. The social class groups whose behavior was significantly different were the lower-lower class, the unemployed, those not in the labor force and the retired. With the exception of the retired households, the households in these four categories probably constitute the major proportion of those in the two lowest educational categories, the only categories whose behavior was found to be significantly different. These results seem to suggest that the two different characteristics, social class and education of the homemaker, do explain much the same portion of the variance in expenditure. It should be noted that the function served by both variables is to help separate out a lower socio-economic group whose behavior differs from the main body of households. Neither characteristic, social class nor education of the homemaker, seems to be particularly significant except insofar as it separates out this deviant group.

Self-Employment of the Head

There is reason to question the reliability of the results obtained in analyzing the effects of self-employment of the head on food expenditures because of the procedures which were used in classifying households. These problems were reported in examining the test of significance for the self-employment characteristic. Because of these problems a final judgment on the significance of self-employment was not made. The regression results will be reported here to permit the reader to exercise his own judgment if he wishes.

A total of 609 of the 3641 households whose expenditure was studied had heads who were classified as self-employed. The results of the regression analysis are presented in Table 3.12. The regression in which the effects of self-employment were estimated also included variables for urbanization.

The self-employed households were found to spend somewhat more than the households whose heads were not self-employed. These results were contrary to those which would have been expected on the basis of the studies of the saving and spending behavior of households with self-employed heads which were discussed in Chapter II. In view of the coding procedures which were used in determining self-employment status, and the unexpected sign and size of the coefficient estimated it appears that this analysis provides no certain basis for a final judgment about either the significance of self-employment or its pattern of effect on food expenditure.

Employment of the Homemaker

The employment of the wife or female head of the household outside the home was not found to influence food expenditures significantly in either a statistical or an economic sense. The results of this test were discussed in Chapter II.

The regression results on which the test of significance was based are given in Table 3.13. Of the 3641 households included in this study, 128 had no wife or female head present or did not report on her employment status.

Households in which the homemaker was employed were found to spend slightly more than those in which she was not. This result conforms to the effect usually assumed for employment outside the home. The difference found was, however, too small to be considered significant.

TABLE 3.12. Regression Results for Self-Employment of the Household Head

Variable	Estimated Coefficient	Computed t	Percent Expenditures Are of Those of Omitted Category
Constant term	-.81474 (.03984)	-20.449	--
Income	+.31468 (.01106)	+28.457	--
SMU	+.59883 (.01276)	+46.938	--
Self-employment of head	+.02229 (.00940)	+ 2.370**	105.3
Rural non-farm	-.05525 (.00661)	- 8.362	--
Rural farm	-.20876 (.01117)	-18.690	--

n = 3641

$\bar{R}^2$ = .55

$\bar{Y}$ = $22.88

Sy^2 = .02757

$\overline{\log Y}$ = 1.29551

TABLE 3.13. Regression Results for Employment of the Homemaker

Variable	Estimated Coefficient	Computed t	Percent Expenditures Are of Those of Omitted Category
Constant term	-.96761 (.04294)	-22.536**	--
Income	+.37624 (.01164)	+32.314**	--
SMU	+.53627 (.01390)	+38.590**	--
Employment of the homemaker	+.01250 (.00694)	+ 1.800	102.9

n = 3513	Sy^2 = .03109
$\bar{R}^2$ = .49	$\overline{\log Y}$ = 1.30097
$\bar{Y}$ = $23.10	

Evidence of Interactions

In the previous sections the S-D variables were treated as if their effects were independent and additive when expressed in logarithms. For example, the effects of region were assumed to be independent of those of urbanization and the effects of the categories of each characteristic were treated as additive logarithms of constants. There is reason to believe, however, that the effects of many of the S-D variables are not independent of the others. There are some indications that the effect of urbanization may differ between regions. We might, for example, hypothesize that the effect of farm residence on food expenditures would differ between the Western region and the North Central region because of the greater amount of home food production which was reported in the North Central region in the 1955 USDA Survey.[1] In order to examine the effects of interaction between urbanization and region, a regression including dummy variables for 11 of the 12 (4 regional x 3 urbanization categories) combinations of urbanization and region was estimated. The urban-North Central category was omitted. The results are given in Table 3.14. The $\bar{R}^2$ obtained was .565 which is only fractionally better than the $\bar{R}^2$ of .564 obtained in a regression (see Table 3.1) which assumed the separate effects of region and urbanization to be independent.

Despite the small difference between the two regression formulations in the percent of the variance of expenditures explained, we find

---

[1] U. S. Department of Agriculture, Food Consumption of Households in the West, "Household Food Consumption Survey 1955," Report No. 5 (Washington: USGPO, 1956), p. 13.

U. S. Department of Agriculture, Food Consumption of Households in the North Central Region," Household Food Consumption Survey 1955," Report No. 3 (Washington: USGPO, 1956), p. 15.

TABLE 3.14. Regression Results for Interaction Categories of Region and Urbanization

Variable	n	Estimated Coefficient	Computed t
Constant term	--	-.74490 (.04097)	-18.181
Income	--	+.29297 (.01124)	+26.063
SMU	--	+.61141 (.01272)	+48.078
Urban-Northeast	738	+.01446 (.00878)	+ 1.647**
Urban-North Central	664	(omitted category)	
Urban-South	556	-.05608 (.00961)	- 5.836**
Urban-West	297	-.01170 (.01144)	- 1.022
Rural non farm-Northeast	229	-.04655 (.01261)	- 3.690**
Rural non farm-North Central	280	-.06791 (.01175)	- 5.779**
Rural non farm-South	324	-.10494 (.01151)	- 9.116**
Rural non farm-West	96	+.00158 (.01796)	+ .088
Rural farm-Northeast	69	-.17029 (.02093)	- 9.136**
Rural farm-North Central	198	-.20574 (.01354)	-15.192**
Rural farm-South	166	.24021 (.01479)	-16.243**
Rural farm-West	24	-.12906 (.03410)	- 3.784**

$n = 3641$

$\bar{R}^2 = .565$

$\bar{Y} = \$22.88$

$S_y^2 =$

$\overline{\log Y} = 1.29551$

that there is evidence that the effects of region do differ between urbanization categories. The results of a series of t tests of the coefficients for each level of urbanization are given in Table 3.15. The tests start with the most negative coefficient in each urbanization and test it against the next most similar coefficient until a significant difference is found. If we group categories with others from which they are not significantly different we would have the following grouping:

Urban categories

(1) Urban-South

(2) Urban-West and Urban-North Central

(3) Urban-Northeast

Rural non-farm categories

(1) Rural non farm-South

(2) Rural non farm-North Central and Rural nonfarm-Northeast

(3) Rural non farm-West

Rural farm

(1) Rural farm-South

(2) Rural farm-North Central and Rural farm-Northeast*

(3) Rural farm-West and Rural farm-Northeast*

* Rural farm-Northeast may be classified in either of these two categories.

We find that residence in the Southern region has much the same effect in each urbanization. In all three urbanizations Southern households spend the least. The effects of Western residence do differ between urbanizations, however. The urban-West category does not behave significantly differently from the urban-North Central category. However, in both the rural urbanizations the Western households do differ significantly from

TABLE 3.15. Tests of the Significance of Differences between Regions at Given Levels of Urbanization

Difference tested	Computed t
Urban	
U-S and U-W	3.75**
U-W and U-NC	1.02
U-W and U-NE	2.29*
U-NC and U-NE	1.65*
Rural Non-farm	
RNF-S and RNF-NC	2.76**
RNF-NC and RNF-NE	1.49
RNF-NC and RNF-W	3.56**
RNF-NE and RNF-W	2.41**
Rural Farm	
RF-S and RF-NC	1.72*
RF-NC and RF-NE	1.55
RF-NC and RF-W	2.17*
RF-NE and RF-W	1.06

those in the North Central region. The effect of Western residence in the rural non-farm and rural farm categories is to make the coefficients estimated far less negative. The effect of Northeastern residence also differs somewhat by urbanization. The urban households spend significantly more than urban-North Central households, but the expenditures of rural Northeastern households do not differ significantly from those of rural North Central households.

These differences suggest that the use of a single set of combined urbanization-regionality categories would give superior results to the use of two separate sets of categories whose effects would be assumed to be strictly additive in logs. Since urbanization and region are the two socio-demographic characteristics which have been found to have the greatest effect on expenditure, it seems desirable to use the most precise techniques possible in dealing with them.

Conclusions About Patterns of Effect

Of the eight S-D variables which have been examined two have been found to be non-significant in effect. Employment of the homemaker was found to be non-significant after the effects of income and SMU are taken into account. Race was found to be non-significant when region as well as income and smu are taken into account.

Although self-employment of the household head was found to be significant, the coefficient estimated after controlling for urbanization calls the reliability of the coding procedures for self-employment into question. The effect of this characteristic should be examined in a future study.

The two most important variables in explaining variance in food expenditure were found to be urbanization and region. These two

variables were found to display some interaction effects, indicating that the effects of the urbanization and the regionality characteristics are not independent and are better treated with the use of combination urbanization-region categories rather than two separate sets of categories, one for urbanization, the other for regionality.

The coefficients estimated for stage in the family life cycle suggested that there are three rather distinct groups of life cycle categories:

(1) (a) young husband and wife households, with and without children present, and (b) households headed by either a male or a female in which children are present.

(2) households headed by a male or female (under age 65) in which no children are present.

(3) older households, including (a) husband and wife households in which the wife is 65 or over and (b) households headed by a male or female age 65 or over.

The dimensions used to delimit the categories were: age, presence of children and marital status. The three groups of life cycle categories differ from each other in at least one of these three dimensions, suggesting that all three should be used in constructing a system of life cycle categories.

An interesting similarity in the results for social class and education of the homemaker has been noted. The effect of social class is significant chiefly because of the behavior of the lower-lower class, the unemployed, those not in the labor force and the retired. The effects of education of the homemaker were significant chiefly because of the behavior of the two categories with elementary school education or below. It was hypothesized that the "deviant" categories of social

class and education of the homemaker contained many of the same households. It appears that the social class and the education variables are useful and significant chiefly because they separate out a group of households of low socio-economic status.

A closer examination of these "deviant" lower socio-economic status households would be useful in order to determine the underlying causes of the differences in behavior observed. Almost certainly, the concentration of older households in these categories explains a part of the differences in their behavior. It may be that controlling for stage in the life cycle would further reduce such significance as social class and education appear to have at this stage in the analysis.

Implications of the Observed Patterns of Effect for the Quantification of the S-D Variables

The use of dummy variables in the foregoing analysis was a deliberate attempt to avoid restrictive assumptions about the pattern of effect of the S-D variables considered. With the use of dummy variables and discrete categories it was possible to avoid the use of scales or continuous variables, such as years of education, which presuppose the effects of a variable to be linear. Had scaled variables been employed in this analysis the results of the tests of significance would likely have been much the same. The regression results would, however, have given us few clues about the real pattern of effects. The effects of social class and education, for example, do have a pattern which, in general, is linear. The use of a linear formulation would, however, have concealed the fact that the only significantly different behavior was at the lower end of the educational scale and among the lower-lower class households and households not classified into regular occupational categories.

The use of scales and continuous variables to represent S-D variables assumes a knowledge of the pattern of effects of S-D characteristics which we do not now have. It is felt that the results obtained in this study demonstrate rather clearly that such scales and variables may obscure more information than they provide and that their use should be avoided at the present stage in the study of the effects of S-D variables.

Alternative economic models of the effects of S-D characteristics will be discussed in the next chapter. The problems of level of measurement and its relationship to various mathematical models will be considered in more detail there.

CHAPTER IV

Alternative Economic Models of the Effects of S-D Variables

Up to this point we have been concerned with the problems of specifying a structural model and of variable specification. We have attempted to determine which S-D variables appear to be significant and determine, on the basis of the pattern of their effects, what particular formulation of the variable appears to be best. We turn now to consider the problems of specifying an economic model and to examine some alternative models. We have, of course, already employed one model of the effects of S-D variables, the regression model on which the results reported in the previous two chapters were based. This model assumed that the presence of a given characteristic would affect expenditure by some constant proportion. A number of other economic models based on other sets of assumptions have been or can be employed in the study of the effects of S-D variables. In this chapter we will attempt to classify these models, examine the assumptions underlying each and point out the strengths and shortcomings of each.

In specifying a model we also specify the level of measurement of the variables which is to be employed. The considerations of Chapter III were largely concerned with the most feasible level of measurement to employ in specifying variables. Or stated differently, Chapter III was concerned with the best way to assign numbers to represent the S-D characteristics. The qualitative nature of most S-D variables tends to dictate the level of measurement to be employed. The level of measurement at which the variables are expressed tends, in turn, to dictate the mathematical model which may be employed, or, at least, to restrict the choice of models.

Since the levels of measurement at which S-D variables are expressed play such a key role in determining the model chosen, we will consider them first and then go on to the consideration of the alternative models themselves. The levels of measurement will be discussed in the next section of this chapter. The alternative economic models of the effects of the S-D variables will be discussed in the following section.

Levels of Measurement and the Models

The Levels of Measurement[1]

On examining the S-D characteristics in which we are interested as a group, we quickly become aware of their diverse nature. Some S-D characteristics can be described by cardinal numbers, while others are numbered ordinally. The observations for still other S-D characteristics do not lend themselves even to ordinal placement. An example of this last group is regionality. We can conceive of an ordinal listing of the regions of the country only if we stipulate some additional criteria to be used as the basis of the ordering; one such criteria might be average temperatures.

Because of this diversity it seems useful to draw on Measurement Theory for a description of the levels of measurement at which variables can be handled.

The simplest form of measurement is the assignment or mapping of individual objects into named or symbolically represented categories.

---

[1]The discussion in this section draws heavily upon S. S. Stevens, "Mathematics, Measurement and Psychophysics," Handbook of Experimental Psychology, S. S. Stevens (ed.), (New York: Wiley, 1951), pp. 1-30 and upon Clyde H. Coombs, "Theory and Methods of Social Measurement," Research Methods in the Behavioral Sciences, Leon Festinger and Daniel Katz (eds.), (New York: Dryden, 1953), pp. 471-88.

This system of categories constitutes a nominal scale. The nominal scale has certain properties which may be regarded more formally as axioms. One such property is the requirement that the relation "equal to" or "not equal to" must hold between all objects in the scale. An object in any given category is equal to any other object in the same category, but is not equal to an object in any other category. In addition, the equality relationship must be transitive and symmetric. The symmetry requirement means that if a = b, then b = a. The transitivity requirement means that if a = b and b = c, then a = c. An example of a nominal scale is the assignment of individuals to the categories "white" and "non-white."

The nominal scale takes account of perceived differences between categories. The differences are not, however, of such a nature as to permit us to say that one class has more or less of some particular attribute than another. In cases where the differences between categories are of a nature which permits us to say that members of one class have more of a particular attribute than those in another, it becomes possible to construct an ordinal scale. This scale arises out of the possibility of ranking members of one category as greater than members of another category in the particular attribute under consideration. In order for an ordinal scale to exist the "greater than" relationship must hold between all pairs of individuals from different categories.

At the next level of measurement individuals are ranked on an interval scale and it is possible not only to state their relative magnitude but also the absolute magnitude of the distance between all pairs of individuals. In order to be able to determine such distances, a distance function which assigns a real number to each pair of elements ordered in the scale is required. For operational purposes all that is

required is a common and constant unit of measurement such that numbers may be assigned to the positions of individuals on the scale. These numbers must be so constructed that arithmetic can be performed on the differences of numbers describing the positions of two separate individuals. The location of the absolute zero for such a scale is arbitrary. Thus the relations between intervals are preserved under the addition of a constant to all scale scores, an operation called "translation." Since the unit of measurement employed is arbitrary, the relation between intervals is preserved under multiplication of all scale values by a constant, i.e., under scalar multiplication. Familiar examples of such a scale are the Centigrade and Fahrenheit measurements of temperature, both of which employ an arbitrarily selected zero point.

The ratio scale is the most commonly used scale in economics. Its familiar uses include the counting of dollars of income and expenditure or units of output. It is a scale of equal intervals, like the interval scale, but its zero point is absolute, rather than of arbitrary choice. The size of the intervals is, however, chosen arbitrarily, just as it is in the case of the interval scale. Because the origin of the scale is absolute, the only legitimate transformation of scale values is by scalar multiplication. The ratio of scale values is preserved under this operation. Other transformations such as the addition of a constant to each scale value are not permissible.

Levels of Measurement Implicit in the Models

As was pointed out at the beginning of the chapter, the qualitative nature of the S-D characteristics often dictates the level of measurement which can be employed. The level of measurement, in turn, tends to dictate the mathematical models which can be employed.

The analysis of variance (AOV) model utilizes the lowest level of

measurement. In order to utilize AOV techniques, we need only place the categories whose effects are under study on the nominal scale. Variables expressed at higher levels of measurement are transformed to the nominal scale when they are employed in AOV. It should be noted here that while the independent variables are expressed in the nominal scale, the "dependent" variable must be expressed in the interval or ratio scale.

In general, only variables on the interval or ratio scales fully qualify for use as regression variables. The results for such variables can be interpreted without qualification. At the lower levels of measurement, certain statistics employed in regression analysis such as means and standard deviation lose meaning. The mean is not a permissable statistic on the ordinal scale, for example, since the mean value of a set of ordinal placements has no real meaning, even though the median value would have meaning.[1]

On occasion, ordinal level variables have been employed in regression analysis. An example would be coding the three levels of urbanization as 1, 2 and 3. The values taken by the variable are such that the assumptions of the least-squares technique are not violated. There is some question about the actual meaning or interpretation of the results, however. While we can treat ordinal level variables as if they were at the interval or ratio scale level, many of the statistics calculated from them can only be interpreted intuitively. Stevens argues that the mean and standard deviation are not permissible statistics on the ordinal level of measurement.[2] The reason for this can be seen when we try to

---

[1] Stevens, p. 25

[2] Stevens, p. 25

interpret the meaning of a mean value of 2.5 for the urbanization variable discussed above when the variable by definition can only take whole number values. While we can produce statistics using ordinal values it is their meaning or validity which is in doubt. The problem is akin to the validity of using football players' numbers as a regression variable. Since the player designations are numbers, we can manipulate them, but since they are on a nominal scale, the results cannot be interpreted in the same fashion as with variables on the ratio or interval scale.

The same cautions are in order in interpreting regression results when dummy variables have been employed. The dummy variable represents membership in categories which are essentially on the nominal scale. Since the variable takes numerical values of zero and one, we can manipulate these values. We must take care in interpreting the statistics which result, however. For example, what is the meaning of a mean value of .57 for a variable, which by definition takes only values of zero and one? We can interpret the .57 value as the proportion of observations falling into the category represented by the "1" value, but should remember that on the nominal scale the means is not a permissible statistic.

Classification of the Alternative Models

A wide variety of economic models has been employed in past studies of the effects of S-D characteristics. It is possible to group these models into three classes according to the assumptions which are made about the effects of the S-D characteristics. The three classes are:

Class I. Models which assume a given S-D characteristic affects only the coefficients of the variable(s) which represent it, influencing the level of expenditure directly through these variables.

Class II. Models which assume a S-D characteristic affects only the coefficient of the income variable, influencing expenditure through its effect on this coefficient. In this class of models it is assumed that the S-D characteristic does not affect the coefficients of other independent variables or the constant term.

Class III. Models which assume the S-D characteristic influences expenditure through its effect on the coefficients of all the independent variables and on the constant term.

Although the alternative models of the effects of S-D characteristics appear on casual examination to be rather similar, there are, in fact, substantial differences between the models in the way in which S-D characteristics are assumed to affect expenditures. The above list may give some appreciation of the differences in assumptions which underlie some of the models which have been employed in empirical studies or have been suggested. The models and their assumptions are examined in detail in the following sections along with examples of their use.

Class I Models

In the group of models which will be labeled "Class I" the assumption is made that a given S-D characteristic affects only those independent variables or constant terms which represent its effects, influencing the level of expenditure directly through these variables. It is assumed that the characteristic has no effect on any of the other variables.

Two major groups of models fall into Class I; they are (a) models which assume the effects of a S-D characteristic are fixed constants and (b) models which assume the effects of a S-D variable may be represented by a continuous variable, properly one on the interval or ratio scale.

Fixed Effects Models

The fixed effects models include both analysis of variance (AOV) models and regression models which employ dummy variables to represent the effects of membership in a particular category. Both models employ the assumption that membership in a given category affects expenditure by a fixed amount.

Analysis of Variance Model. The AOV model which can be employed for both tests of significance and for the estimation of the effects of membership in given categories has been labeled "Model I" or the "Fixed Effects Model."[1] This model assumes that the effects of all factors which influence expenditure can be represented by constants. The model assumes that the expenditure of household t can be represented as the sum of the mean expenditure of all households, M, plus the additive constants a_i and b_j representing the effect of membership in category i of household characteristic A and category j of characteristic B, plus a random error u_{ijt}. The model is:

$$Y_{ijt} = M + a_i + b_j + u_{ijt} \qquad (4.1)$$

where $i = 1, \ldots I$

$j = 1, \ldots J$

where $M = \dfrac{\sum_j \sum_i Y_{ij\cdot}}{I \cdot J}$

and $Y_{ij\cdot}$ is the arithmetic mean of observations in cell ij, and where the u_{ijt} are independently distributed $N(0, \sigma^2)$. The model is determinate

---

[1] The differences in the assumptions of the two basic AOV models are discussed in detail in Snedecor, pp. 257-62.

only when the restriction that $\sum a_i = \sum b_j = 0$ is imposed.[1]

This model assumes that the effects of characteristic A and B are additive and independent of each other, i.e., that there is no interaction.

Regression Model with Dummy Variables. The regression model which employs dummy variables to represent the effects of membership in the categories of a given S-D characteristic assumes that the effects of the categories can be represented by fixed constants, while the effects of other characteristics are represented by continuous variables.

A regression model in which independent variable X_t represents the value of the continuous variable income for household t, and dummy variable Z_{it} represents membership in category i of a given S-D characteristic is:

$$Y_{it} = bX_t + \sum_i c_i Z_{it} + u_{it} \qquad (4.2)$$

where $i = (1, \ldots I)$

$t = (1, \ldots T)$

and Y_{it} is the observed value of expenditure of household t, which falls in S-D category i,

X_t is the observed value for household t of income

Z_{it} = 1 when household t is in category i of a given characteristic

= 0 otherwise

u_{it} are independently distributed $N(0, \sigma^2)$.

It was this model, which was employed by Martin David to explain the consumption of automobiles in lower-income households (categoryl).[2] The

---

[1] Henry Scheffe, The Analysis of Variance (New York: Wiley, 1959), pp. 60-62.

[2] M. H. David, Family Composition and Consumption ("Contributions to Economic Analysis," No. 25; Amsterdam: North-Holland, 1962), pp. 85-93.

actual model employed by David was:

$$Y_{ilt} = a_1 X_t + \sum_i b_{il} Z_{ilt} + u_{ilt} \qquad (4.3)$$

where $i = 1, \ldots 12$, where $j = 1$,

and where Y_{ilt} is total value of cars owned by household t, which is in family composition category i, income level category l

X_t is disposable income for household t

$Z_{ilt} = 1$ when household t falls in category il

$= 0$ otherwise

It can be demonstrated that the model employed by Jean Crockett to estimate the constant term for particular city classes was basically of this type also. Mrs. Crockett estimated regressions for each of 9 city classes, but restricted the regression coefficients of the income and family size variables to values estimated previously.[1] The regression coefficients of both the income variable (X) family size variable (N) were thus the same in all 9 equations with only the constant terms differing between equations. The basic model thus was:

$$\log Y_{it} = \log a_i + .4835 \log X_t + .3346 \log N_t + \log u_{it} \qquad (4.4)$$

where $(i = 1, \ldots 9)$

$(t = 1, \ldots T)$

and where Y_{it} is the food expenditure of household t in city class category i. Since logarithms are used in this model, membership in a category i affects expenditure by a fixed proportion rather than a fixed dollar amount.

---

[1] Crockett, pp. 303-304.

Continuous Variable Models

The fixed effect models in Class I assume that the effects of membership in the categories of a S-D variable can be represented by a set of fixed constants. In contrast to these models, the continuous variable models assume a straight line or linear relationship between the continuous variable representing a S-D characteristic and the expenditure variable.

An example of such a model where age of the household head A is assumed to affect expenditure in a linear fashion is:

$$Y_t = a + bX_t + CA_t + u_t \qquad (4.5)$$

where Y_t is the expenditure of household t,

X_t is the value of the income variable for household t,

A_t is the age of the head of household t.

The use of a model which assumes a linear relationship between the variable representing a S-D characteristic and food expenditure may lead to serious errors arising from the improper specification of the S-D variable. Before specifying a S-D variable in this fashion, the researcher would do well to assure himself that its effects are, in fact, linear. The effects of many S-D variables may have non-linear patterns of effect or no patterns at all. The general absence of patterns in the effects of the S-D categories presented in Chapter III illustrates the danger of using a continuous variable without some prior investigation.

Continuous variables representing the effects of a S-D characteristic have been employed in a number of studies. For many of these studies it is not altogether clear that the propriety of the assumption of linear effects was investigated thoroughly before a linear variable was employed.

The use of age of the household head as a linear variable in explaining household food expenditure is clearly contraindicated by Jean Crockett's findings that the effects of age of head on levels of food expenditure are approximately parabolic.[1]

Class II Models

The models grouped in Class II assume that the effects of a given S-D characteristic on expenditure occur through its effect on the coefficient of the regression of expenditure on income. The coefficient of the income variable is assumed to differ among the categories of a S-D characteristic, with each S-D category affecting expenditure through its effect on the income-expenditure relationship.

An example of such a regression model is:

$$Y_{it} = a + \sum_{i} b_i X_{it} + c W_t + u_{it} \qquad (4.6)$$

where $t = 1, \ldots T$ and $i = 1, \ldots I$,

and where

Y_{it} = the observed value of the dependent variable for household t, which falls in S-D category i,

X_{it} = the observed value of the income variable when household t is in S-D category i.

= 0 when household t is not in S-D category i.

W_t = the observed value for household t of some other independent variable (income excluded).

In this model the b_i differ among the I S-D categories. The coefficient c is not, however, affected by the S-D category into which a household falls, nor is the constant term affected.

---

[1] Crockett, pp. 307-308.

Class III Models

The models grouped in Class III assume the S-D characteristic under study affects the regression coefficients of all the independent variables and also the constant term. Regression coefficients and constant terms are assumed to differ between S-D categories. The S-D category affects expenditure through its effects on both the regression coefficients and constant term.

Models of this type are usually expressed as a set of I equations, each one representing one of the I categories of a given S-D characteristic:

$$\begin{aligned} Y_{1t} &= a_1 + \sum_k b_{1k} X_{1kt} + u_{1t} \qquad (4.7) \\ &\cdots \\ Y_{it} &= a_i + \sum_k b_{ik} X_{ikt} + u_{it} \\ &\cdots \\ Y_{It} &= a_I + \sum_k b_{Ik} X_{Ikt} + u_{It} \end{aligned}$$

where $i = 1, \ldots I$, and $k = 1, \ldots K$

and where Y_{it} is the expenditure of household t, which is classified in category i of the S-D characteristic under study, and X_{ikt} is the value for household t of the k-th independent variable. The constant terms a_i of this model differ between equations as do the b_{ik}.

The same assumptions could be incorporated in a single equation model, if desired. The model suggested by Daniel Suits to deal with interaction between a qualitative variable and a continuous variable is of this type.[1] The model suggested by Suits includes a separate constant term for each S-D category and a separate regression coefficient for the continuous variable for each S-D category. Such a model could be stated:

---

[1] Suits, p. 550.

$$Y_{it} = \sum_i a_i Z_{it} + \sum_i b_i X_{it} + u_{it} \qquad (4.8)$$

where $i = 1, \ldots I$

and

Z_{it} = 1 when household t falls in category i

= 0 otherwise

X_{it} = observed value of income (or other independent variable) when household t is in category i

= 0 otherwise

The multiple equation version of this model has been used widely. It was employed by Jean Crockett in her study of the effects of age of the household head on food expenditures.[1] Separate regressions were estimated for each of the seven age of head categories employed. The model used was:

$$\log Y_{it} = a_i + b_i \log X_{it} + c_i \log N_{it} + u_{it} \qquad (4.9)$$

where Y_{it} is food expenditure (excluding alcoholic beverages) for home consumption of household t which falls in S-D category i, X_{it} is family income after taxes and N_{it} is family size. On examining the results, Mrs. Crockett concluded that the differences in the b_i and C_i between categories were small and proceeded to re-estimate the seven regressions using a model in which only the constant term (a_i) differed between age categories. This second model employs the assumptions of Class I.

A Class III model was employed by the author in his study of the effects of household age-sex composition on the income-expenditure elasticity for food.[2] The author estimated a separate regression for

---

[1] Crockett, pp. 306-307.

[2] Herrmann, op. cit.

each of the 26 age-sex composition types which he employed.

The Class III model has also been employed in the study of non-food expenditures. One of the models employed by Martin David in his study of family consumption of automobiles falls into Class III.[1] David estimated eight separate regressions, each one representing a particular combination of income level and family composition. The model employed to explain total value of cars owned (Y) was:

$$Y_{ijt} = a_{ij} Z_{ijt} + b_{ij} X_{ijt} + u_{ijt} \qquad (4.10)$$

where $i = 4, 5, 6, 10$

$j = 1, 2$

$t = 1, \ldots 2646$

and

Y_{ijt} = observed total value of cars owned for family t, which is classified in income category i and family composition category j

X_{ijt} = observed value of disposable income for family t

Z_{ijt} = 1 when household t falls in income category i and family composition category j

= 0 otherwise

David concluded that the b_{ij} did not differ among households in the lower income category (j = 1) and proceeded to estimate a single regression for these households with dummy variables representing membership in the different family composition categories and a single income variable. This second model falls into Class I.

---

[1] David, pp. 85-93.

The above examples suggest that the Class III models are the most general of the three classes which have been discussed. The S-D characteristic under study is assumed to affect all the independent variables and constant terms. On examining the results the use of one of the other models may be indicated. This was the procedure followed by both Mrs. Crockett and David in the studies cited. After testing the relevance of the most general model, they proceeded to a simpler model with its more restrictive assumptions about the effects of S-D characteristics.

Some Conclusions about Economic Model Specification

It is felt that it would be useful to develop a single regression model which would estimate the magnitude of the effects of the significant economic and socio-demographic variables. The resultant regression would be useful in explaining the level of average expenditures in the past and in forecasting future levels. The most suitable form of regression probably would be one which could utilize information on the marginal distribution of each of the variables and would not require the use of joint distributions of characteristics, since such joint distributions are often difficult or impossible to obtain.

Pragmatic considerations thus seem to indicate that a Class I model would be the best choice if independent and additive effects can be assumed. The use of Class II models for predictive purposes would necessitate projections of the joint distribution of income and the S-D characteristic under study. The use of a Class III model would necessitate projections of the joint distributions of the S-D characteristic under consideration and each of the independent variables employed.

A Suggested Combined Model

On the basis of our tests of significance and the examination of alternative models of the effects of S-D models, it appears possible to sketch the general form of an economic model including all the significant S-D variables. This combined model should, on the basis of the tests discussed previously, include S-D variables for region, urbanization and stage in the family life cycle. An additional variable taking account of the atypical behavior of households in lower socio-economic categories seen in the examination of the results for social class and education of the homemaker might also be desirable.

The most easily manageable model including a number of S-D variables would be one which assumes that the regression coefficient of the income and meals variables are independent of the effects of the S-D characteristics and that each of the S-D characteristics affects expenditure by a constant amount or proportion. The proposed model would have the following characteristics:

1. Double-log relationship between income and expenditure.
2. Include a meals variable, probably total meals served, with a logarithmic transformation.
3. Include region and urbanization as separate variables or as a single set of combined region-urbanization categories.
4. Include a life cycle stage variable; a tentative set of categories would be
   (a) Husband-wife households (wife under age 65) plus households with male or female head (under 65) with children present,
   (b) Households with male or female head (under age 65), with no children present,

(c) Households with male or female head (over age 65), children may be present plus husband-wife households (wife over 65), no children present.

5. The model probably also should include a variable to take account of the atypical behavior of lower socio-economic groups found in the estimates of the effect of social class and education of the homemaker.

In listing the characteristics of a proposed model the problems remaining in the development of such a model become clearer. Questions remain about whether the urbanization and region characteristics should be expressed as separate variables or as a single combined variable, and about how the deviant behavior of the lower socio-economic categories should be handled. These two problems appear to be the principal ones remaining before a combined model including S-D characteristics can be constructed.

It also might be useful to investigate alternative formulations of the meals and family composition variable. Experience with the SMU variable in this study indicates that the adjustment of total meals served for family composition gives only negligible improvement in the $\bar{R}^2$ obtained over that obtained with unadjusted total meals served. The results with total meals served and a separate family composition variable (e.g., the sum of the adult equivalent scale values for members of the family) might be worth investigating.

Multi-collinearity in a Combined Model

It would appear on casual examination that certain problems might arise from inter-correlations of the S-D variables and the other independent variables when all are included in a single regression equation. Multi-collinearity would increase the standard errors of the regression

coefficients and reduce the accuracy of the estimates of the effects of individual characteristics. The predictive usefulness of the coefficients outside of the actual regression estimated would be in doubt.

On closer examination it appears that the multi-collinearity problem may not be so serious as might be anticipated. There is some independent evidence that the intercorrelation of income and socio-economic variables would not be so high as might be expected. In his study of income distribution, F. Gerard Adams employed an equation of the general form

log Income = f (Age, Education, Occupation, Region, Community Size, Part of the Year Worked).[1]

Six categories of age were employed along with a squared variable to take account of the parabolic effect of age on income. Two categories of education were employed, three categories of occupation, two of region, three of community size (degree of urbanization) and two of part of year worked. With this predictive equation for income, Adams obtained a coefficient of multiple correlation of .66. This suggests that there would be rather substantial correlation between the socio-demographic variables and the income variable included in the proposed Combined Model. However, the intercorrelation probably would be below the figure obtained by Adams, as three of the variables employed by Adams, i.e., occupation, part of year worked and education would, it now appears, be replaced by a single variable devised to take account of membership in the lower socio-economic categories.

---

[1] F. Gerard Adams, "The Size of Individual Incomes: Socio-economic Variables and Chance Variation," Review of Economics and Statistics, Vol. 40 (November, 1958), pp. 390-98.

It would, however, be desirable to determine the extent of correlation between income and the S-D variables to be employed in the combined model before such a model actually is constructed. This might be done by estimating $\bar{R}^2$ for a regression with income as the dependent variable and the S-D variables to be employed in the combined model as the independent variables. It is felt that the actual correlation will not be so great as to impair the usefulness of the combined model.

CHAPTER V

Conclusions and Implications

At this point it is clear that certain further problems remain before an economic model including socio-economic and demographic (S-D) variables can be specified in full. It is felt, however, that important progress has been made toward such a model. Urbanization, region and stage in the family life cycle have been identified as three key S-D variables which must be included in any such model. The absence of any clear patterns in the effects of the S-D variables suggests that dummy variables should be used to represent S-D characteristics rather than continuous variables on the interval or ratio scale. The examination of alternative economic models indicated that a model specifying the S-D variables in such a way that their effects are additive constants (or multiplicative constants when logarithms are used) would provide the most useful estimates of effects.

Certain additional conclusions appear when the study is examined from an overall standpoint. Such conclusions, based on the results for the study as a whole, are presented in the first portion of this chapter.

The estimates of the effects of the S-D characteristics which have been obtained permit us to make some preliminary calculation of the effects of changes in population composition on average food expenditures. With such calculations, we can determine tentatively whether the differences in expenditure behavior between categories are large enough for changes in population composition to affect aggregate food expenditure. If these preliminary calculations suggest that changes in population composition do have an important effect on aggregate expenditure then the goal of a single economic model including S-D variables

clearly should be pursued to a final conclusion. Calculations of the effects of changes in the distribution of the population by urbanization and region on average food expenditure are presented in the second section of this chapter.

Conclusions

Comments on the Use of Socio-demographic Variables

Alvin Gouldner has noted that the applied social sciences have tended to adopt concepts from social science theory and employ these concepts in applied research without much consideration of hypotheses which might be suggested by the theories in which the concepts were embedded.[1] This observation seems to apply to the use of socio-economic and demographic variables in economics and marketing as much as to other areas of applied social science.

Such variables as social class and stage in the family life cycle have been employed without much examination of the theory underlying them. This neglect of the theory underlying these socio-economic and demographic variables may, perhaps, be excused in the initial stages of research. Preliminary attempts to determine the significance of an S-D characteristic are often rawly empirical. As research in this area becomes refined it seems desirable to return and examine the theory underlying significant socio-demographic variables. In this way the hypothesis about the manner in which a socio-demographic variable affects expenditure behavior can be refined and sub-hypotheses about the behavior of particular categories can be developed. A re-examination of basic theory taken along with the empirical results

---

[1] Alvin W. Gouldner, "Explorations in Applied Social Science," Social Problems (January, 1956), reprinted in Marketing and the Behavioral Sciences, Perry Bliss (ed.), (Boston: Allyn and Bacon, 1963), pp.6-7.

already obtained should provide a basis for improving the category divisions employed and explaining such differences as have been found between the categories in the initial analyses.

Now that some initial results concerning the effects of social class and life cycle stage are available, it appears that the next step in research should be an attempt to interpret these results in the light of available theory. Such an interpretation clearly is needed before further work is done with either of these variables. Similar interpretations of the observed effects of regionality and urbanization are also needed. The available theory underlying these two variables is, however, little developed in contrast to the theory involving social class and stage in the family life cycle.

What is needed then, is continuing interaction between applied social science and the basic social science theories whose concepts have been employed. It appears that a re-examination of applied social science results in the light of existing basic theory would be as fruitful as the initial adoption of theoretical concepts for use in applied social science research.

The Effect of the Use of SMU and S-D Variables on Income Elasticity Estimates

The inclusion of additional independent variables, such as Standard Meal Units and S-D variables, may be expected to affect the regression coefficient of the income variable to the extent that these additional variables are correlated with income. The inclusion of the SMU variable can be expected to have lowered the coefficient estimated for the income variable. The simple correlation of log income with log SMU was found to be .23. This correlation is the net effect of two divergent

influences. One influence is the overall positive correlation of income and household size, which suggests a positive correlation of income and the total number of meals eaten by members of a household. The second influence is the negative correlation of income and the number of meals eaten at home within a given household size type.

The coefficient estimated for the income variable in a regression including only income and SMU as independent variables was .378. The effect of the addition of S-D variables can be seen when we compare this figure to the coefficients of income obtained in regressions which include S-D variables. The income coefficients in regressions which included S-D variables were, without exception, lower than .378. The estimate in the regression which included region and urbanization variables, in addition to those for income and SMU, was .293. The estimate in the regression which included social class and urbanization variables was .281, while the estimate in the regression including life cycle, urbanization and regional variables was .275. The lowest estimate obtained was .257, which occurred in the regression which included variables for urbanization, region and the education of the homemaker. Since the above coefficients were estimated in regressions of the double-log form they all may be interpreted as income elasticities.

These effects of the inclusion of SMU and S-D variables on the estimate of income elasticity are similar to those observed by Jean Crockett in her examination of the 1950 B.L.S. data. Mrs. Crockett notes:

> There is evidence that large families, families living in large cities or suburbs in the North, and families with middle-aged heads, all tend to spend more on food than other families at the same income level. Since these families are relatively more frequent at high than at low incomes, the failure to hold these factors constant may be expected to lead to unduly high estimates of income elasticity.[1]

---

[1] Crockett, p. 295.

The failure to hold the effects of family size, the number of meals eaten at home and S-D characteristics constant does, it appears, result in cross-sectional income elasticity estimates which are too high. By holding at least some of these variables constant, we have obtained cross-sectional estimates of income elasticity as low as .257 in this study.

The Importance of S-D Variables in Explaining Food Expenditure

The relative contributions of the significant S-D variables to the explanation of the variance of food expenditure were discussed in Chapter II. The order of their relative importance was:

urbanization
region
life cycle stage
education of the homemaker
social class

It is estimated these five S-D variables explain an additional 7.5 percent of the variance of food expenditures beyond the 49.5 percent explained by the income and SMU variables. The addition of S-D variables, thus, increases the percent of the variance in expenditure explained by about 15 percent. Such an improvement appears to be of sufficient magnitude to be of interest and concern.

Implications

The estimates of the effects of the S-D characteristics which have been obtained permit a preliminary evaluation of the impact of changes in population composition on average expenditures for food. While estimates from a single regression including variables representing all the relevant S-D characteristics would be desirable, the presently

available estimates of effects do permit some tentative conclusions about the effects of changes in population composition.

Estimates of the Effects of Recent Changes in Population Composition

The tests of significance and regression analyses indicate that urbanization, region and stage in the family life cycle all have substantial effect on household food expenditure behavior. In addition, it was noted certain categories of education of the homemaker and social class behaved in a fashion which differed markedly from the other categories. The categories whose behavior was atypical were those with little education and low socio-economic status.

Fairly substantial differences in expenditure behavior have been found between the categories of the significant characteristics. These differences between categories will, over time, have little effect on the total demand for food unless changes occur in the distribution of households among these categories. We can evaluate the possible effects of changes in the distribution of households among regions, urbanization categories and stages in the family life cycle by considering the magnitude of the changes in these distributions in recent years.

The changes in the distribution of households between the three categories of urbanization can be seen in Table 5.1. Between 1950 and 1960 the percentage of the households in the urban and rural farm categories declined, while the percentage in the rural nonfarm category increased rather substantially. The effect of this change in distribution can be estimated by using the estimates of relative levels of food expenditures along with distribution information to estimate the weighted average expenditure (see Table 5.1). In 1950, average food expenditure for all households was 92.9 percent of the average urben expenditure. In 1960, the average household food expenditure was 93.7

TABLE 5.1. Changes in the Distribution of Households by Urbanization and Resultant Changes in Expenditure, 1950 to 1960

	Percentage of Households		Expenditures as a Percentage of Urban Average[d]	Weighted Expenditure	
	1950[a]	1960[b]		1950	1960
Urban	66.1	63.6	100.0	66.1	63.6
Rural non-farm	19.9	26.6[c]	89.0	17.7	23.7
Rural farm	14.0	9.8[c]	64.8	9.1	6.4
Weighted Average Expenditure as a Percentage of Urban Average	--	--	--	92.9	93.7

[a]U. S. Bureau of the Census, Current Population Reports, Population Characteristics, Series P-20, No. 33 (February 12, 1951), p. 13. Based on 1950 definition of household, does not include Alaska or Hawaii.

[b]U. S. Bureau of the Census, Current Population Reports, Population Characteristics, Series P-20, No. 103 (July 6, 1960), pp. 1-2, Based on 1960 definition of household, includes Alaska and Hawaii.

[c]Based on 1950 definition of farm.

[d]See Table 3.1.

percent of the average urban expenditure. The change in the distribution of households between urbanization categories thus moved average U. S. expenditure .8 percent closer to the urban average expenditures. On the basis of a mean expenditure for food at home in urban households in the 1955 U.S.D.A. study of $24.23, the .8 percent change would amount to 11 cents, in 1955 dollars.[1] This change, although small, on an individual household basis, becomes of greater importance when it is aggregated over a total of more than 50 million U. S. households and calculated on a yearly basis. Based on 53.0 million households in 1960, the change in average expenditure amounts to some $303 million a year.[2]

The changes in the distribution of households between regions can be seen in Table 5.2. The proportion of households located in the North East, the North Central region and the South decreased between 1950 and 1960. The proportion of households located in the West increased by 2 percent during this period. The effects of this shift in the distribution of households by region are estimated in Table 5.2. In 1950, the U. S. average household food expenditure was 98.7 percent of the average expenditure in the North Central region. In 1960 the U. S. average expenditure was 98.8 percent of the North Central average. The estimated net effect of the shifts in the distribution of households by region was to increase average household food expenditure by 2.4 cents per week. This amounts to an increase of about $66 million a year when

---

[1] U. S. Department of Agriculture, Food Consumption of Households in the United States, p. 11

[2] U. S. Bureau of the Census, Statistical Abstract of the United States: 1961: (82nd ed.; Washington, USGPO, 1961), p. 40.

TABLE 5.2. Changes in the Distribution of Households by Region and Resultant Changes in Expenditure, 1950 to 1960

	Percentage of Households[a]		Expenditures as a Percentage of North Central average[d]	Weighted Expenditure	
	1950[b]	1960[c]		1950	1960
North East	26.1	25.5	104.7	27.3	26.7
North Central	30.2	29.0	100.0	30.2	29.0
South	29.4	29.2	89.7	26.4	26.2
West	14.3	16.3	103.4	14.8	16.9
Weighted average expenditure as a percentage of North Central average	--	--	--	98.7	98.8

[a]U. S. Bureau of the Census, Statistical Abstract of the United States: 1961 (82nd ed.; Washington, USGPO, 1961), p. 40.

[b]Includes Alaska and Hawaii. Based on 1950 definition of household.

[c]Based on 1960 definition of household.

[d]See Table 3.1.

summed over all households.[1]

The extent of changes in the distribution of households by age of the household head can be seen in Table 5.3. Between 1950 and 1960 the proportion of households with younger heads declined somewhat while the proportion with older heads increased somewhat. The proportion of households with heads in middle-aged categories stayed fairly stable. The categories employed in this study in the examination of the effects of life cycle stage do not match those employed in the Census. It is, therefore, impossible to estimate effect of changes in the distribution of households by age of head with the data now available. It would be expected that the 3.1 percent increase in households with heads age 65 and over would lower the average level of expenditure, since households with heads over age 65 were found to spend only about 81 percent of the amount spent by couples, with the oldest child between age 6 and 15 (the omitted category).[2]

It appears that the current changes in population composition, although gradual, have been large enough to produce some changes in average levels of expenditure. The change in urbanization can be expected to have increased average expenditure, while the shift in the

---

[1] Jean Crockett has estimated that changes in the distribution of households between combined region-urbanization-race categories will have negligible effect on average food expenditures between 1950 and 1970, since the distributional effects appear to cancel each other out. See Jean Crockett, "Population Change and the Demand for Food," Demographic and Economic Change in Developed Countries, National Bureau of Economic Research ("Special Conference Series," No. 11; Princeton: Princeton University Press, 1960), pp. 472-75.

[2] Jean Crockett has estimated that changes in the distribution of households by age of head may be expected to have significant negative effect on average food expenditure between 1950 and 1970, in contrast to the negligible effects of shifts in distribution between combined region-urbanization-race categories. See Jean Crockett, "Population Change and the Demand for Food," pp. 469-72, 475.

TABLE 5.3. Changes in the Distribution of Households by Age of Head, 1950 to 1960

	Percent of Households	
	1950[a]	1960[b]
Under 25 years old	4.9	4.8
25 to 34 years old	20.6	18.4
35 to 44 years old	22.5	22.0
45 to 54 years old	20.0	20.6
55 to 64 years old	17.2	16.3
65 years and over	14.8	17.9

[a]U. S. Bureau of the Census, "Marital Status and Household Characteristics: March 1950," Current Population Reports, Population Characteristics, Series P-20, No. 33 (February 12, 1951), p. 15. Based on 1950 definition of household, does not include Alaska or Hawaii.

[b]U. S. Bureau of the Census, "Household and Family Characteristics: March 1960," Current Population Reports, Population Characteristics, Series P-20, No. 106 (January 9, 1961), p. 13. Based on 1960 definition of household, includes Alaska and Hawaii.

distribution between regions can be expected to have had a lesser effect in the same direction. The changes in the distribution by age of head can be expected to have reduced average expenditure somewhat. All in all, it appears that the net effect of changes over the ten-year period, 1950 to 1960, in these three socio-demographic variables has been large enough to be of some real significance. The longer-term effect undoubtedly has been far larger, in view of the substantial shifts in the distribution of the population by urbanization, region and age, which have occurred in the past 50 years.

Further Studies Needed

Some of the problems remaining in developing a combined model were discussed in Chapter IV. These problems include:

(1) examination of the effects of replacing the SMU variable with two separate variables, one for total meals, the other representing household composition,

(2) determining whether region and urbanization should be included as separate variables or as a single combined variable,

(3) determining a way to take account of the atypical behavior of households in the lower socio-economic categories,

(4) determine the probable extent of multi-collinearity between the variables to be included in the model.

It appears that once these problems are resolved a combined economic model can be constructed.

Another problem requiring further study is the significance of the effects of self-employment of the household head. This question could not be resolved in this study because of the inadequacy of the information available on self-employment status. In view of the substantial

effect self-employment has on other areas of consumer expenditure, it would be desirable to evaluate its effect on food expenditures.

An additional problem which requires further study is the atypical behavior of the households in the lower socio-economic categories in this study. It would be useful to discover the key causal factor at work, or at least the socio-demographic characteristic which appears to explain the deviant behavior of this group the best. It is suspected that the behavior of this group is not so homogeneous as has been supposed. It includes a variety of household types, in a wide variety of situations. The group includes those who are unable to work because of age, the disabled, the unemployed and those employed at unskilled labor. There is every reason to believe that despite possible similarities in their levels of income, the family of an aged pensioner and the family of a young unskilled laborer would spend in a very different manner. Differences in the behavior of different types of low-income households have important implications for the development of food stamp plans and the distribution of welfare payments. Because of these implications, the assumption of homogeneous expenditure behavior in low-income households should be tested.

The three classes of models which have been set forth in this study would appear to be a useful classification to employ when studying the effects of other types of household characteristics on expenditure. They appear to be a useful way of thinking about the manner in which certain psychological variables, such as attitudes and expectations, may affect expenditure. It would be interesting and useful to study the applicability of these models with psychological variables and other variables describing household characteristics besides the socio-demographic variables which have been employed in this study.

APPENDIX I

The Data Employed

The 1955 Household Food Consumption Survey Sample[1]

The data which are utilized in this study were drawn from the sample of the U. S. Department of Agriculture's 1955 Household Food Consumption Survey. The purpose of the U.S.D.A. study was to collect current information on patterns of food consumption, expenditures, dietary levels and household food practices. The study includes data from 6,060 households in which at least one member of the household ate 10 or more meals from household food supplies in the preceding week. Of the 6,060 households from which complete interview schedules were taken, 4,556 were in the basic national sample. The basic national sample was supplemented by interviews of 1,504 farm-operator households, which were selected in the same manner as the basic national sample, but were drawn from rural areas only. The interviews were collected in the months of April, May and June, 1955.

Only households in which one member ate 10 or more meals in the preceding week were considered eligible. The 4 percent of the population which lives in non-housekeeping units, such as rooming houses, institutions and prisons, were thus omitted from the study. Of the dwelling units selected for the basic sample, 5,551 were found to be occupied. Of these, 5,140 were found to be eligible for inclusion in this study since they had served 10 or more meals to at least one person in the

---

[1] This appendix summarizes the description of the survey and the sampling procedures employed given in U. S. Department of Agriculture, Food Consumption of Households in the United States, pp. 1-3, 186-95.

previous week. Complete schedules were collected for 4,556 of these eligible households in the basic sample. The other eligible households had no one at home or refused to supply the desired information. Only 93 percent of the occupied dwelling units selected in the basic sample were found to be eligible, 89 percent of these eligible households actually participated in the survey by supplying the information requested.

The survey employed the "recall-list" method and used specially trained interviewers to help respondents recall the quantities of food used during the week and the amounts paid for purchased items. The information collected was on food consumed or used up, rather than on purchases. In addition to this information on the quantities used at home for all food items in the previous week, information was collected on the number of meals eaten at home and away by each individual in the household and on expenditures for food eaten away from home.

The basic sample was a national, self-weighting, area, probability sample. It was drawn by first separating metropolitan areas with a population of over 50,000 from the non-metropolitan areas. The 168 metropolitan areas were divided into five size classes on the basis of their population in the 1950 Census. Each of the 8 metropolitan areas in the over 2 million population size class was regarded as a separate stratum. The other metropolitan areas were grouped into 3 to 8 geographic areas within each of the remaining four size classes. Each of these groups of metropolitan areas had a population of approximately 2 million and were made up, with few exceptions, of varying combinations of states within the four Census regions, Northeast, North Central, South and West. This grouping produced 32 metropolitan area strata, 8 of which consisted of a single major metropolitan area, with the remaining 24 strata each consisting of a group of smaller metropolitan areas of the same approximate size, all located within the same region.

In each of the 8 major metropolitan area strata, cities were classified by size. One city was selected from each size stratum, for a total of 23 cities. A total of 22 cities was selected from the other 24 metropolitan area strata. The 45 cities selected were chosen with a probability proportional to their 1950 population.

The 605 cities with populations between 10,000 and 50,000 outside the standard metropolitan areas were grouped within Census regions into 15 strata, averaging about 85 thousand people per strata. One city was selected from each of these 15 strata with probability proportional to its 1950 population. These 15 strata for cities outside standard metropolitan areas and the 45 strata for cities within standard metropolitan areas provided a total of 60 cities for the national sample.

In the 32 standard metropolitan areas, the 81 counties (excluding cities of over 10,000) were grouped into 33 strata. The remaining 2,697 counties outside the standard metropolitan areas were grouped into 82 strata of contiguous economic areas, each having about 650,000 persons. One county was selected from each of these 115 county strata. A total of 60 cities and 115 counties were thus selected for the sample.

Within the cities, sample segments were selected by numbering all the segments set up within the city and drawing every n-th one with probability proportional to the number of dwelling units in the segment. Sample segments were selected randomly in each county from a listing of Census enumeration districts which excluded places of 10,000 or more population. One smaller segment was selected with equal probability from each of these enumeration districts. A total of 2,000 segments was selected from the 60 cities and 115 counties in the sample, of these 1,527 were allocated to the basic sample.

The dwelling units to be visited within each segment were selected

from lists made up in a specified geographic order. A random number was used to select the first sample dwelling unit to be interviewed, it and every n-th succeeding unit were then designated for interview. The number "n" was determined by dividing the expected number of dwelling units in the segments by the expected number of units to be visited. Each segment was expected to yield 3 interviews. No substitutions were provided for households which were not eligible to participate, or which were absent or refused to participate. In urban areas, interviewers were required to make as many as four return visits if necessary, in order to make original contact.

The survey sample may be appraised by comparing it with Census data. The distribution of households by number of persons, the proportion of non-white households and the proportion owning television sets were found to be much alike.

About 11 percent of the eligible households failed to participate in the survey by supplying the requested information. The resulting non-response bias was considered to be small on the basis of comparisons of the answers to certain questions obtained from those classified as non-respondents with those obtained from participating households, and the comparatively small proportion of non-respondents. The proportion of non-participants to participants was smaller than average in the South and larger in the North Central region. The non-participating households had fewer husband-wife households and fewer household members. The members of the non-participating households included fewer children and more adults over 50. A higher proportion of the housewives in the non-participating households were employed.

Selection of Household Observations from the HFCS Sample

The data employed in this study were selected from the basic national sample portion of the 1955 Household Food Consumption Survey. The observations of the supplementary farm sample were not employed. Observations were taken only from the basic national sample in order that this study would be, insofar as possible, representative and nationwide in scope.

Households in the basic national sample which did not report income could not be employed in this study and were separated from the portion of the sample which was used. The households which did not report income fall into two categories: (1) households which were not asked for their 1955 income because they were only recently established, and (2) households which participated in the survey, answering all questions except the question about income. The resultant bias has not been investigated. Studies on the non-reporting group in previous food consumption surveys indicated that while the nonreporting group did differ from the rest of the sample, they were not different enough or numerous enough to distort the total results.[1]

Households in the basic sample reporting annual incomes below $750 were also omitted from the group selected for this study. Relatively few households fell in this category. It was noted, however, that a major proportion of the households in this excluded category reported food expenditures at a weekly rate which would have been in excess of their reported annual income.

In addition to the income requirements which have been discussed, two compositional requirements were stipulated for households selected

---

[1] Faith Clark, Janet Murray, Gertrude S. Weiss and Evelyn Grossman, p. 182.

from the basic national sample for use in this study. The first stipulation required that the household consist only of a primary economic family. "Economic family" was defined in the 1955 HFCS as "a person living alone or a group of persons who lived together and drew from a common fund for major items of expense."[1] If more than one such economic family was present in a household the one most closely connected with maintaining the dwelling unit was regarded as the primary economic family. The income information collected was that for the primary economic family. No information was collected on the income of secondary economic families in cases where they were present. The expenditure information for food used at home, however, included expenditures for food served to secondary economic families, boarders and guests as well as that served to the primary economic family. The use of the income of the primary economic family as the determinant of this expenditure would have understated the income level actually influencing food expenditure. It was decided, therefore, to exclude from this study households in which secondary economic families or boarders were present.

The second compositional stipulation required that the number reported as eating 10 or more meals in the household during the survey week equal the number in the primary economic family. This requirement was necessary because it was planned to employ a variable based in part on the age-sex composition of the primary economic family (the standard meal units variable) as one of the independent variables. It was expected that the Standard Meal Units variable would correct for differences between households in composition and in the number of meals served.

---

[1] U. S. Department of Agriculture, Food Consumption of Households in the United States, p. 193.

The compositional correction would, however, have been incorrect or incomplete for those households in which the number of regular meal eaters, i.e., the number eating ten or more meals during the survey week, differed from the number in the primary economic family. For this reason households in the basic sample in which the number eating ten or more meals during the week did not equal the number in the primary economic family were not selected for inclusion in this study.

After the imposition of these requirements there were 3641 households of the 4,556 participating households included in the 1955 U.S.D.A. Survey which could be included in this study.

Handling of Missing Information

The problems raised by non-response to survey questions concerning income were discussed in the previous section. However, we often find other items of information, which are missing from completed and otherwise acceptable survey schedules. The omissions arise from the respondent's refusal or inability to answer certain questions and from interviewer errors in completing the survey schedule.

John Lansing and A. T. Eapen suggest that survey schedules with key missing information be completed by assigning responses where omissions have occurred.[1] The process utilized to complete schedules at the University of Michigan Survey Research Center relies on information from previous surveys. Omissions are filled-in on the basis of this information. The effect of assignment of responses on the final distributions obtained does not appear to be large, however. Klein

---

[1] J. B. Lansing and A. T. Eapen, "Dealing with Missing Information in Surveys," Journal of Marketing, Vol. 24 (October, 1959), pp. 25-26.

notes that, "parallel calculations for non-assigned and assigned-plus-nonassigned cases show nearly identical results."[1]

It was decided to omit interviews with missing information from this study when necessary. Interviews which did not provide income information were omitted altogether, as was mentioned in the previous section. Interviews which omitted other items were omitted only when the omitted information was to be employed as a variable. Thus, an interview which omitted information on race was not employed in any of the analyses of the significance and effects of race, but was employed in all other analyses. The number of observations, thus, varies somewhat between analyses. If Klein's experience with Survey Research Center data is applicable, the effect of omitting these interviews with missing information rather than assigning responses is probably small.

Two of the analyses deal with characteristics of the homemaker, her education and her employment status. Some of the respondent households did not contain a wife or female head. These households were omitted in the analysis of the effects of the education and the employment of the homemaker along with households which did not supply this information.

---

[1] Lawrence K. Klein, "Statistical Estimation of Economic Relations from Survey Data," Contributions of Survey Methods to Economics, Lawrence R. Klein (ed.), (New York: Columbia University Press, 1954), p. 208.

APPENDIX II

The Standard Meal Units Variable

Previous work with the data employed in this study found a high degree of variability between households in the number of meals eaten at home.[1] The number of meals eaten at home was found to differ a great deal even between households of the same size and age-sex composition. A total meals served variable was employed in that study to control for this source of variation in expenditure.

The use of total meals served as an explanatory variable, however, suggests that every meal was of the same size, e.g., that the portions of infants and adults were of the same size. It was felt that some correction was required to standardize the total number of meals served for differences in portion sizes. The result would be a meals variable expressed in standardized units. This variable will be labeled the Standard Meal Units (SMU) variable.

In order to obtain the SMU variable, consumption equivalence scales for 13 age-sex categories were first estimated. The scales were based on the estimated cost of one week's food for each age-sex type under the U. S. Department of Agriculture's Moderate Cost Food Plan. The male, age 20-49, category was used as a base category. The cost of one week's food under the Moderate Cost Food Plan was estimated at $8.80 (1957 prices) for this category. The cost of one week's food for each of the other age-sex types was divided by $8.80 to obtain costs relative to the base category. The estimated cost of one week's food and the consumption equivalence coefficients of each of the age-sex categories are presented in Table II.1.

---

[1] Herrmann, pp. 71-2.

The sum of the scale coefficients of the members of each household was then obtained. This value divided by the number of members of the household was used as a coefficient of adjustment by which the total number of meals served was multiplied. The result is the number of standardized meal units served in the particular household during the survey week. The coefficient of adjustment thus adjusts total meals served upward or downward into SMU on the basis of probable portion sizes for the various age-sex categories. The scales are based, of course, on recommended nutritional allowances rather than objective data. The use of this information in the form of relative scales, however, probably mitigates some of the errors arising from its normative nature.

Further refinement of the standardization process would, of course, be possible. It might, for example, be desirable to take into consideration the number of meals eaten by members of each age-sex category and to take account of differences between breakfasts, lunches and dinners. Such highly refined corrections do not seem to be merited on several grounds. The basic survey data with which we are working undoubtedly contains both sampling and response errors of sufficient magnitude as to make over-refined adjustment techniques unwarranted. In addition, a comparison of the $\bar{R}^2$ obtained for the SMU variable as compared to that with a total meals served variable suggests that even the use of an adjustment to SMU may not be worth the expense and effort involved.

TABLE III.1. Cost of One Week's Food Under the Moderate-Cost Food Plan and Consumption Equivalence Scale Coefficients for Thirteen Age-sex Categories

	Cost of One Week's Food[a]	Scale Coefficient
Under 1 year	$3.50	.40
1 to 5 years	$4.65	.53
6 to 11 years	$6.54	.74
12 to 14 years, female	$8.00	.91
12 to 14 years, male	$8.75	.99
15 to 19 years, female	$8.05	.91
15 to 19 years, male	$10.25	1.16
20 to 49 years, female	$7.03	.80
20 to 49 years, male	$8.80	1.00
50 to 64 years, female	$6.58	.75
50 to 64 years, male	$8.17	.93
65 and over, female	$6.50	.74
65 and over, male	$8.00	.91

[a]Prices are for June, 1957. See Eloise Cofer, "Family Food Budgets, Revised 1957," U. S. Department of Agriculture, Family Economics Review, October, 1957 (Washington: USDA), p. 11.

APPENDIX III

The Social Class Variable

Discussion of the Procedure Employed

Households in this study were assigned to social class categories on the basis of the occupation of their head. The assignment process involved three steps: (1) the assignment of the household head to one of the four hundred detailed occupational classifications employed in the 1950 Census on the basis of occupation as reported to the survey interviewer, (2) the assignment of a Socio-economic Index (SI) score to the household on the basis of the detailed occupational classification of its head, (3) the assignment of households to a social class category on the basis of the SI score.

The Socioeconomic Index employed is one developed by Otis Dudley Duncan and builds from the findings of a study of occupational prestige conducted in 1947 by Paul K. Hatt and Cecil C. North for the National Opinion Research Center.[1]

Duncan's SI seemed particularly suited to the needs of the present study and the information which was available on which to base a social class assignment. Information on the contents of the household living room used in scales of social status such as Chapin's "living-room scale"[2] was not, of course, available. Scales such as Warner's Index of Status Characteristics[3] are designed for use in individual

---

[1] Otis Dudley Duncan, Chap. VI: "A Socioeconomic Index for all Occupations" and Chap. VII: "Properties and Characteristics of the Socioeconomic Index," in Albert J. Reiss, Jr., Otis Dudley Duncan, Paul K. Hatt and Cecil C. North, Occupations and Social Status (Glencoe, Illinois: Free Press, 1961).

[2] F. Stuart Chapin, Contemporary American Institutions (New York: Harper, 1935), Chap. 29.

[3] W. Lloyd Warner, Marchia Meeker and Kenneth Eels, Social Class in America: The Evaluation of Status (New York: Harper, 1960).

localities, but are not suitable for studies covering a wide area. The Socioeconomic Index is, however, applicable to nationwide samples and SI scores can be assigned on the basis of occupational information alone. Since this study required an index which was nationwide in scope and could be used with the data available, the SI seemed particularly suitable. It must be emphasized that the Socioeconomic Index should be regarded as an index of social class rather than as a synonym for social class. Duncan regards the SI as merely one possible measure of social stratification and recognized that its applicability would be governed by the nature and objectives of the project under consideration.[1]

A detailed discussion of the procedure employed in assigning households to social class categories is given in the following sections.

Interpretation of Occupational Information

The occupation of the household head was obtained from each household interviewed in the 1955 Household Food Consumption Survey. This information had not been utilized previous to this study and had not been coded in any way. In using the SI it was necessary first to assign the household head's reported occupation to one of the four hundred detailed occupational classifications employed in the 1950 Census. SI scores were available for each one of these categories.

Occupational information as recorded by the interviewer was reviewed and the head's occupation was assigned to one of the detailed occupational classifications on the basis of this information. In cases where the terminology used to describe the head's occupation was

---

[1] Otis Dudley Duncan, "A Socioeconomic Index for All Occupations," in Reiss et al., pp. 139-40.

unfamiliar, The Alphabetical Index of Occupations and Industries was employed to ascertain the conventional occupational classification.[1]

In cases where two occupations were reported for the head, he was assigned to the classification which appeared to be his principal occupation. The few households which did not report the occupation of the head were assigned to the "Occupation Not Reported" classification. This classification was employed in the 1950 Census and the SI provides a score for the category.

Many households reported occupational information about the head which indicated that he (or she) was not gainfully employed. Some of these individuals were unemployed or retired. Others had no occupation because they were unemployable or had never chosen to enter the labor force. Because SI scores were available only for regular occupational categories it was necessary to assign households whose heads were not gainfully employed to three special groups. The three groups were: "Unemployed," "Retired," and "Not in the Labor Force." Students, housewives, the disabled and households whose head was in prison were assigned to the "Not in the Labor Force" category. It was planned to use these three groups along with the social class groups which were to be derived.

The Socio-Economic Index Scores

The goal of the researchers who prepared the SI was to produce an index of occupational prestige and socio-economic status which could be used in the study of social stratification. It should be emphasized at this point that the index is one of occupational prestige rather than

---

[1] U. S. Bureau of the Census, 1960 Census of Population, Alphabetical Index of Occupations and Industries (Rev. ed.; Washington: USGPO, 1960).

of social class. It is useful in the study of social class structure insofar as differences in occupational status determine differences in social strata.

The SI was developed from the findings of the National Opinion Research Center (NORC) study published in 1947 which presented prestige ratings of 90 occupations. These ratings were the result of a mass survey of opinion on the prestige of different occupations. The Socio-economic Index Scores estimated by Duncan are an attempt to determine prestige ratings of occupations which were not included in the original NORC study.

A review of the results of the NORC study showed that the chief determinants of prestige appeared to be the educational attainment of the members of an occupation and the income distribution within the occupation. The weights of these two factors in determining the prestige of an occupation were estimated from a regression in which the dependent variable was the percent of the respondents rating an occupation as "good" or "excellent" in the NORC study. The percent in an occupation with education through high school or beyond and the percent with incomes over $3,500 for 1949 were employed as the independent variables. The observations employed were those for 45 of the 90 occupations included in the NORC study. The occupations omitted were highly specialized ones for which the required Census data were not available.

The regression coefficients estimated for income and education were then employed with the Census data on the income and education for each of the cateogries in the detailed occupational classification to estimate the percent of "good" and "excellent" responses which would have been expected had the category been included in the NORC survey.

The resultant estimation of the percent of respondents giving an occupation an "excellent" or "good" rating is the Socio-economic Index score of that occupation. Scores for all of the occupational categories in the Census detailed occupational classification were estimated in this manner.

One of the chief problems of the Socio-economic Index is the use of cash income and education as the only independent variables. Because scores are based solely on these two variables, the scores for occupations which receive part of their income in kind or as perquisites are lower than might otherwise be expected. The effect of the choice of determining variables is particularly apparent in the scores for farmers and clergymen. Duncan felt that although the scale might be biased in such cases, any adjustments would be arbitrary.[1]

This author made no changes in scoring, with one exception, for the same reason. An exception was made in the case of the category "Members of the Armed Forces." The SI lumps all members of the Armed Forces into the single category, which has a score toward the lower end of the scale. Because of the heterogeneity of the group included in the category, the members of the Armed Forces were divided into three categories and scored separately for the purposes of this study. Officers were given the score of "Managers and Officials, not elsewhere classified," and non-commissioned officers were given the same score as "Craftsmen, not elsewhere classified." Other enlisted men were given the score included in the Index for "Members of the Armed Forces."

It should be noted that the SI was designed chiefly to determine the occupational prestige of males in particular occupational classifications.

---

[1] Duncan, "A Socioeconomic Index for All Occupations," in Reiss et al., pp. 131-32.

The validity of the SI for women employed in particular occupational classifications has not been investigated.[1] However, for the purposes of this study, SI scores were assigned to households on the basis of the reported occupation of the head, without regard to whether the head was a male or a female.

The Assignment to Social Class Groups

The final step in the process was the assignment of each household included in this study to a social class category on the basis of its SI score. As has been pointed out, Duncan's Socio-economic Index is an index of occupational prestige, not of social class status. Because occupation has been found to be related closely to social class status by students of social stratification and has been regarded by Warner as a principal determinant of social status, this author has chosen to use it as the basis for assigning the households included in this study to the social class categories which will be employed.[2]

The Number of Social Class Groups

Before deciding upon the number of social class categories which were to be used in this study it was first necessary to make some determination of the number of social classes which can be considered to exist in present-day American society. In his pioneering studies of social class, W. Lloyd Warner found that all segments of society classified their fellow citizens into class groups and that the rankings and divisions between classes were fairly consistent despite the social class

---

[1] Duncan, "Properties and Characteristics of the Socio-Economic Index," in Reiss et al., p. 150.

[2] Warner, Meeker and Eels, pp. 176-85.

status of the observer.[1] Since all segments of society perceived social classes to exist and because of the consistency in ranking and assignment of social class status which he found, Warner felt that social classes do exist and could be treated as an objective fact.

Researchers in the area of social class have found the number of social classes recognized to vary between the communities.[2] Warner found six classes in his Yankee City study of an old New England city,[3] while he found only five in his Jonesville study of a midwestern community.[4] The smaller number of classes in the Jonesville study was due to the absence of a division of the Upper Class into Upper-Upper and Lower-Upper groups. This was a result of the fact that Jonesville had no old and established aristocracy which segregated itself from the newly rich. A study, by August Hollingshead, of New Haven found five classes,[5] while his Elmtown study of the same city as Warner's Jonesville found five classes.[6]

Despite the differences in the number of classes perceived to exist, all the studies found that the respondents showed a high degree of consistency in ranking and in the fixing of cutting points between class groups. To avoid further problems about the number of classes

---

[1] Warner, Meeker and Eels, pp. 68-70.

[2] A discussion of the number of classes is given in Joseph A. Kahl, The American Class Structure (New York: Rinehart, 1957), pp. 19-44.

[3] W. Lloyd Warner and Paul S. Lunt, The Social Life of a Modern Community, Vol. I of Yankee City Series (New Haven: Yale University Press, 1941), p. 88.

[4] W. Lloyd Warner, American Life - Dream and Reality (Rev. ed.; Chicago: University of Chicago Press, 1962), pp. 82-83.

[5] August B. Hollingshead and Frederick C. Redlich, "Social Stratification and Psychiatric Disorders," American Sociological Review, Vol. 18 (April, 1953), pp. 164-67.

[6] August B. Hollingshead, Elmtown's Youth (New York: Wiley, 1949), pp. 83-84.

which exist, Warner's six class system was accepted. Only four social class groups will be used, however. It was expected that the number of households included in this study which would fall into the Upper-Upper or Lower-Upper class would be very small. These two groups were therefore combined with the Upper-Middle class. The four class groups employed were: (1) the combined Upper-Upper, Lower-Upper, Upper-Middle group, (2) Lower-Middle, (3) Upper-Lower, and (4) Lower-Lower.

Additional Non-class Categories

In addition to the four social class groups employed it was necessary to provide categories for households to whom no SI score could be assigned. These categories were households whose head was reported as (1) retired, (2) unemployed, or (3) not in the labor force. Such households were assigned directly to the applicable non-class category.

Selection of Cutting Points

After settling upon the number of social class groups into which the households in this study were to be assigned, it was necessary to establish the SI score levels which would be the cutting points for assignment into different social classes. It was desired to select cutting points which would assign households to social class groups in a distribution resembling those found by Hollingshead in his New Haven study and Warner in his Yankee City study. These two studies are the only ones which have attempted to describe the social class distribution of a complete community. Warner's"Yankee City" is a New England town of 17,000, while Hollingshead's New Haven study deals with a city of 250,000. There is little evidence on which to base any judgment as to how closely these distributions may resemble the social class distribution

in the U. S. as a whole. In the absence of any other information these two studies will be used for guideline purposes in this study.

It was necessary to take account of the probable under-representation of the Lower-Lower class, since this group typically has been under-represented in most past household surveys. This problem and those already discussed suggested that only an approximation to the class distributions of Warner and Hollingshead should be sought. This approach was also indicated by the approximate nature of the scores of the Socio-economic Index.

It was decided that cutting points should be set so that class membership followed the general pattern:

Class 1: Important managers and officials, professionals.

Class 2: White collar office workers and skilled craftsmen in jobs requiring substantial skill and special training.

Class 3: Craftsmen in building trades and semi-skilled operatives.

Class 4: Unskilled laborers.

On the basis of this general pattern, cutting points were then selected:[1]

Class 1: SI scores of 65 and above.

Class 2: SI scores 35-64

Class 3: SI scores 10-34

Class 4: SI scores 0-9

---

[1] Duncan suggests that the optimum cutting point between white-collar and manual occupations would be 38.5. This cutting point minimizes misclassifications of occupations when they are divided into a manual and a white-collar category on the basis of their index scores. See Otis Dudley Duncan, "Properties and Characteristics of the Socio-Economic Index," in Reiss et. al., p. 159. The choice of 34.5 in this study approximates this division. The choice was made before the Duncan study was available in its entirety.

The SI scores of some typical occupational categories in each of the class categories are given in Table III.1.

The distribution of households which resulted when the selected cutting points were employed is given in Table III.2, along with the distributions found by Warner for "Yankee City" and by Hollingshead for New Haven. The distribution for this study can be seen to resemble, in general, those found by Warner and by Hollingshead. The households in this study which could not be classified into social class categories probably would fall mostly into the two lowest social class groups. If they were added to these groups the distribution in this study would resemble those found by Warner and Hollingshead even more closely.

The percent assigned to the combined upper-upper, lower-upper and upper-middle class category in this study was quite similar to the combined percentages in the Warner and Hollingshead studies. The percent assigned to the lower-middle class lies between the Warner and Hollingshead percentages as does the percent assigned to the upper-lower class. The percent assigned to the lower-lower class differs rather markedly from the Warner and Hollingshead studies. This difference is probably explained by (1) the large number of households which could not be assigned SI scores, but which probably should fall in the two lower class categories, and (2) the under-representation of lower income households in this study. The first problem has already been discussed. Under-representation in this study of low income households could arise from several sources. One possible cause of under-representation is the requirements which were set for inclusion in the sample of the 1955 Household Food Consumption Survey. A disproportionate number of lower-lower class households were probably excluded from the survey sample because they were not housekeeping households, i.e.,

TABLE III.1. Socio-Economic Index Scores of Typical Occupations Included in the Four Social Class Categories Employed

Category	Occupation	Score[a]
1 (scores 65 and above)	Dentist	96
	Physician	92
	College faculty member	84
	Optometrist	79
	Insurance salesman	66
	State government official	66
2 (scores 35-64)	Hardware store manager	64
	Photoengraver	64
	Librarian	60
	Bank teller	52
	Retail sales clerk	39
	Radio repairman	36
3 (scores 10-34)	Plumber	34
	Machinist	33
	Stonecutter	25
	Baker	22
	Farmer (owner or tenant)	14
	Taxi driver	10
4 (scores 0-9)	Janitor	9
	Carwasher	8
	Farm wage laborer	6
	Porter	4
	Lumberman	4
	Laborer in cigarette factory	0

[a] Reiss et al., pp. 263-75.

TABLE III.2. Distribution of Households Between Social Class Categories in This and Previous Studies

Class Category	Percent of Population		
	Warner-Yankee City[a]	Hollingshead-New Haven[b]	This study-Nationwide
Upper-Upper	1.4	3	11
Lower-Upper	1.6		
Upper-Middle	10.2	8	
Lower-Middle	28.1	22	25
Upper-Lower	32.6	46	42
Lower-Lower	25.2	18	9
Unknown	.9	3	-
Unclassified: Retired, Unemployed or Not in Labor Force			13

[a] W. Lloyd Warner and Paul S. Lunt, Social Life in a Modern Community, Vol. I of Yankee City Series (New Haven: Yale University Press, 1941), p. 88

[b] August B. Hollingshead and Frederick C. Redlich, "Social Stratification and Psychiatric Disorders," American Sociological Review, Vol. 18 (April, 1953), p. 167.

ones in which one member ate 10 or more meals in the week preceding the interview. Those living in rooming houses, hotels and institutions are excluded by this requirement. The exclusion of all households in the survey sample which reported annual incomes below $750 undoubtedly also reduced the number of lower-lower class households in this study.

Those who object to the use of the SI as a basis for assignments to social class categories may prefer to regard the categories employed as SI score categories rather than as social class categories. Social class labels have been given to the SI score categories used in this study because of the familiarity of social class terminology. The results can, however, be interpreted in terms of SI scores, if desired.

BIBLIOGRAPHY

Public Documents:Reports

Clark, Faith, Janet Murray, Gertrude S. Weiss and Evelyn Grossman. Food Consumption of Urban Families in the United States - With an Appraisal of Methods of Analysis, U. S. Department of Agriculture, Agriculture Information Bulletin No. 132 (Washington: USGPO, 1954).

U. S. Bureau of the Census. 1960 Census of Population, Alphabetical Index of Occupations and Industries (Rev. ed., Washington, USGPO, 1960).

U. S. Bureau of the Census. Statistical Abstract of the United States: 1961 (82nd ed.), (Washington: USGPO, 1961).

U. S. Department of Agriculture. Food Consumption of Households in the United States, "Household Food Consumption Survey 1955," Report No. 1 (Washington: USGPO, 1956).

U. S. Department of Agriculture. Food Consumption of Households in the North Central Region, "Household Food Consumption Survey 1955," Report No. 3 (Washington: USGPO, 1956).

U. S. Department of Agriculture. Food Consumption of Households in the West, "Household Food Consumption Survey 1955," Report No. 5. (Washington: USGPO, 1956).

U. S. Department of Agriculture. Food Consumption and Dietary Levels of Households as Related to Age of Homemaker, United States - By Region, Household Food Consumption Survey 1955 Report No. 14 (Washington: USGPO, 1959).

Public Documents:Articles

Cofer, Eloise. "Family Food Budgets, Revised 1957," U. S. Department of Agriculture, Family Economics Review (October, 1957), pp. 1-11.

Humes, Helen M. "Family Food Expenditures, 1947 and 1948," U. S. Department of Labor, Monthly Labor Review, Vol. 66 (June, 1949), pp. 621-630.

Murray, Janet. "Food Consumption and Dietary Levels of Households With Employed and Non-employed Homemakers," U. S. Department of Agriculture, Family Economics Review (June, 1960), pp. 9-31.

U. S. Bureau of the Census. "Marital Status and Household Characteristics: March 1950," Current Population Reports, Population Characteristics, Series P-20, No. 33 (February 12, 1951).

U. S. Bureau of the Census. "Household and Family Characteristics: March 1960," Current Population Reports, Population Characteristics Series P-20, No. 106 (January 9, 1961).

U. S. Bureau of the Census. "Educational Attainment: March 1962," Current Population Reports, Population Characteristics, Series P-20, No. 121 (February 7, 1963).

Books

Anderson, R. L. and T. A. Bancroft. Statistical Theory in Research (New York: McGraw-Hill, 1952).

Burk, Marguerite C. Influences of Economic and Social Factors on U. S. Food Consumption (Minneapolis: Burgess, 1961).

Chapin, F. Stuart. Contemporary American Institutions (New York: Harper, 1935).

Cochran, W. G. Sampling Techniques (New York: Wiley, 1953).

Cochrane, Willard G. and Carolyn Shaw Bell. The Economics of Consumption (New York: McGraw-Hill, 1956).

Croxton, Frederick E. and Dudley J. Cowden. Applied General Statistics (2nd ed.; Englewood Cliffs, N. J.:Prentice-Hall, 1956).

David, Martin H. Family Composition and Consumption ("Contributions to Economic Analysis," No. 25; Amsterdam: North-Holland Publishing Co., 1962).

Friend, Irwin and Robert Jones (eds.). Proceedings of the Conference on Consumption and Saving (Philadelphia: University of Pennsylvania, 1960).

Heady, Earl O. and John L. Dillon. Agricultural Production Functions (Ames: Iowa State University Press, 1961).

Hollingshead, August B. Elmtown's Youth (New York: Wiley, 1949).

Kahl, Joseph A. The American Class Structure (New York: Rinehart, 1957).

LIFE. LIFE Study of Consumer Expenditures, Vol. I. (New York: Time, Inc., 1957).

National Bureau of Economic Research. Demographic and Economic Change in Developed Countries ("Special Conference Series," Vol. 11; Princeton: Princeton University Press).

Prais, S. J. and H. S. Houthakker. The Analysis of Family Budgets ("University of Cambridge, Department of Applied Economics Monographs," No. 4; Cambridge: Cambridge University Press, 1955).

Rainwater, Lee, Richard P. Coleman and Gerald Handel. Workingman's Wife: Her Personality, World and Life Style (New York: Oceana, 1959).

Reiss, Albert J., Jr., Otis Dudley Duncan, Paul K. Hatt and Cecil C. North. Occupations and Social Status (New York: Free Press, 1961).

Reissman, Leonard. Class in American Society (Glencoe, Ill.: Free Press, 1959).

Scheffe, Henry. The Analysis of Variance (New York: Wiley, 1959).

Snedecor, George W. Statistical Methods - Applied to Experiments in Agriculture and Biology (5th ed.; Ames, Iowa: Iowa State College Press, 1959).

Theil, H. Economic Forecasts and Policy ("Contributions to Economic Analysis," No. 15; Amsterdam: North-Holland Publishing Co., 1961).

Warner, W. Lloyd and Paul S. Lunt. Social Life of a Modern Community, Vol. I of Yankee City Series (New Haven: Yale University Press, 1941).

Warner, W. Lloyd. American Life - Dream and Reality (Rev. ed.; Chicago: University of Chicago Press, 1962).

Warner, W. Lloyd, Marchia Meeker and Kenneth Eells. Social Class in America: The Evaluation of Status (New York: Harper, 1960).

Articles and Journals

Adams, F. Gerard. "The Size of Individual Incomes: Socio-Economic Variables and Chance Variation," Review of Economics and Statistics, Vol. 40 (November, 1958), pp. 390-98.

Barton, S. G. "The Life Cycle and Buying Patterns," Consumer Behavior, Vol. II: The Life Cycle and Consumer Behavior, Lincoln H. Clark (ed.), (New York: New York University Press, 1955), pp. 53-57.

Coombs, Clyde H. "Theory and Methods of Social Measurement," Research Methods in the Behavioral Sciences, Leon Festinger and Daniel Katz (eds.), (New York: Dryden, 1953).

Crockett, Jean. "Demand Relationships for Food," Proceedings of the Conference on Consumption and Saving, Vol. I, Irwin Friend and Robert Jones (eds.), (Philadelphia: University of Pennsylvania, 1960), pp. 293-310.

Crockett, Jean A. "Population Change and the Demand for Food," Demographic and Economic Change in Developed Countries, National Bureau of Economic Research ("Special Conference Series," No. 11; Princeton: Princeton University Press, 1960), pp. 457-483.

Gouldner, Alvin W. "Explorations in Applied Social Science," Social Problems (January, 1956), reprinted in Marketing and the Behavioral Sciences, Perry Bliss (ed.), (Boston: Allyn and Bacon, 1963), pp. 4-24.

Hollingshead, August B. and Frederick C. Redlich. "Social Stratification and Psychiatric Disorders," American Sociological Review, Vol. 18 (April, 1953), pp. 161-9.

Kahl, Joseph A. and James A. Davis. "A Comparison of Indexes of Socio-economic Status," American Sociological Review, Vol. 20 (June, 1955) pp. 317-25.

Kish, Leslie. "Confidence Intervals for Clustered Samples," American Sociological Review, Vol. 22 (April, 1957) pp. 154-165.

Klein, Lawrence R. "Statistical Estimation of Economic Relations from Survey Data," Contributions of Survey Methods to Economics, Lawrence R. Klein (ed.), (New York: Columbia University Press, 1954), pp. 189-240.

Klein, L. R. "Entrepreneurial Saving," Proceedings of the Conference on Consumption and Saving, Vol. II, Irwin Friend and Robert Jones (ed.), (Philadelphia: University of Pennsylvania, 1960), pp. 297-335.

Lansing, John and James Morgan. "Consumer Finances over the Life Cycle," in Consumer Behavior, Vol. II: The Life Cycle and Consumer Behavior, Lincoln Clark (ed.), (New York: New York University Press, 1955), pp. 36-53.

Lansing, John B. and Leslie Kish. "Family Life Cycle as an Independent Variable," American Sociological Review, Vol. 22 (October, 1957), pp. 512-19.

Lansing, J. B. and A. T. Eapen. "Dealing with Missing Information in Surveys," Journal of Marketing Vol. 24 (October, 1959), pp. 21-27.

Martineau, Pierre. "Social Classes and Spending Behavior," Journal of Marketing, Vol. 23 (October, 1958), pp. 121-30.

Miller, Donald L. "The Life Cycle and the Impact of Advertising," Consumer Behavior, Vol. II: The Life Cycle and Consumer Behavior, Lincoln H. Clark (ed.), (New York: New York University Press, 1955), pp. 61-65.

Morgan, James N. and John A. Sonquist. "Problems in the Analysis of Survey Data, and a Proposal," Journal of the American Statistical Association, Vol. 58 (June, 1963), pp. 415-34.

Selvin, Hanan C. "A Critique of Tests of Significance in Survey Research," American Sociological Review, Vol. 22 (October, 1957), pp. 519-27.

Stevens, S. S. "Mathematics, Measurement and Psychophysics," Handbook of Experimental Psychology, S. S. Stevens (ed.), (New York: Wiley, 1951).

Stigler, George J. "The Early History of Empirical Studies of Consumer Behavior," Journal of Political Economy, Vol. 62 (April, 1954), pp. 95-113.

Suits, Daniel B. "Use of Dummy Variables in Regression Equations," Journal of the American Statistical Association, Vol. 52 (December 1957), pp. 548-51.

Woodbury, Robert M. "Economic Consumption Scales and their Uses," Journal of the American Statistical Association, Vol. 39 (December, 1944), pp. 455-68.

Zwick, Charles. "Demographic Variation: Its Impact on Consumer Behavior," Review of Economics and Statistics, Vol. 39 (November, 1957), pp. 451-56.

Bulletins and Reports

Bilkey, Warren J. The Basic Relationships in Consumer Expenditure Behavior ("Harvard Studies in Marketing Farm Products," No. 4-H; Cambridge: Harvard University, 1951).

University of Pennsylvania. Study of Consumer Expenditures, Incomes and Savings, Vol. III: Expenditures for Food, Beverages and Tobacco (Philadelphia: University of Pennsylvania, 1956).

Unpublished Material

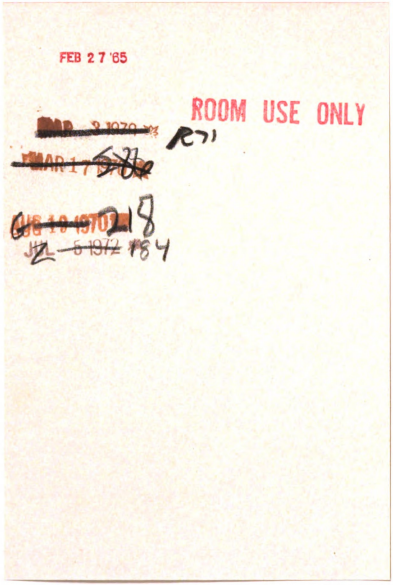
Gustafson, Robert L. "The Use and Interpretation of 'Dummy Variables' in Regressions," Note of January 22, 1962 (Revised), Department of Agricultural Economics, Michigan State University, East Lansing (Mimeographed).

Herrmann, Robert O. "An Investigation of Differences in Income Elasticities of Demand for Food in Households of Differing Size and Composition," (unpublished Master's thesis, Department of Agricultural Economics, Michigan State University, 1961).

Moss, Thomas N. "Some Relationships of Selected Socio-Economic Factors to Food Consumption and Expenditures, Lansing, Spring, 1950," (unpublished Ph.D. dissertation, Department of Agricultural Economics, Michigan State College, 1952).

Sparks, Willard Robert. "Estimates of the Demand for Food from Consumer Panel Data," (unpublished Ph.D. dissertation, Department of Agricultural Economics, Michigan State University, 1961).

ROOM USE ONLY



MICHIGAN STATE UNIV. LIBRARIES
3129310170792 9