MARKET SEGMENTATION THROUGH ZIP CODE ANALYSIS

A CASE STUDY

By

PAUL JEFFREY SCANLON

Bachelor of Science

Oklahoma State University

Stillwater, Oklahoma

1969

Submitted to the Graduate Faculty of the
Department of Administrative Sciences of
the College of Business Administration
of the Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF BUSINESS ADMINISTRATION
May, 1974

#254

MARKET SEGMENTATION THROUGH ZIP CODE ANALYSIS

A CASE STUDY

Report Approved:

Stephen J. Miller

Report Adviser

Head, Department of Administrative Sciences

PREFACE

This study is concerned with a market segmentation analysis of the publics served by the Oklahoma Tuberculosis Association. The primary objective is to identify the demographic characteristics of individuals or geographic aggregations of individuals that explain and allow prediction of success in the solicitation for donations. A stepwise multiple regression analysis is used to examine the data from an Internal Revenue Service Report divided by zip code for the State of Oklahoma and from the donor records from the Association.

The author wishes to express his sincere appreciation to his major advisor, Dr. Stephen J. Miller, for his guidance and assistance throughout this study. Appreciation is also expressed to Administrative Sciences faculty members, Dr. William M. Kincaid, and Dr. Ralph Catalanello for their assistance.

Thanks must also be given to Mr. Ralph Morgan and his staff at the Oklahoma City office of the Tuberculosis Association, and to Mr. George Martin and his staff at the Tulsa - Lakes area office of the Association, for making this study possible and for their valuable assistance during the study. Thanks are also extended to Dena Meenen and her staff at the University Computing Center for programming assistance.

Finally, special appreciation is expressed to my wife, Cindy, who typed the many drafts and final copy of this manuscript, and without whose sacrifice and understanding this paper would not have been completed.

TABLE OF CONTENTS

Chapter		Page
I.	INTRODUCTION AND PURPOSE	
	Introduction	1
	Objectives of the Study	3
II.	STATEMENT OF THE PROBLEM	5
III.	SURVEY OF THE LITERATURE	
	Marketing's Application to Non-Business Organizations	9
	Use of Demographic and Socioeconomic Data In Marketing Research	12
IV.	RESEARCH METHODOLOGY	
	Determining the Sample	17
	Collection of the Data	26
	Method of Analysis	34
V.	ANALYSIS OF DATA AND INTERPRETATION OF RESULTS	
	Independent Variables	38
	Dependent Variables	41
	Results of the Regression Analysis	43
VI.	SUMMARY AND FINDINGS	
	Summary	59
	The Research Findings	60
VII.	DECISION IMPLICATIONS AND SUGGESTIONS FOR FUTURE STUDY	
	Decision Implications	62
	Suggestions for Further Study	67
APPENDIZES		
	Appendix A - O.T.B.R.D.A. Marketing Study	72
	Appendix B - Model Application	74

LIST OF TABLES

Table		Page
I.	Oklahoma 5-Digit Zip Codes, Categorized by Population Levels and Summary Totals for each Level	17
II.	Summary of 5-Digit Zip Codes Selected for the Study	18-23
III.	Counties Appearing in Sample and Number of Times Appeared	24-26
IV.	Explanation of the Independent Variables	39-41
V.	Explanation of the Dependent Variables	43
VI.	Introductory Statistics - Stepwise Regression Analysis	44-45
VII.	Correlation Matrix	46-47
VIII.	Summary Statistics - Model I	49
IX.	Summary Statistics - Model II	52
X.	Summary Statistics - Model III	54

LIST OF FIGURES

Figure		Page
1	Data from Internal Revenue Service Tax Report	30
2	Donor File Card - Oklahoma Tuberculosis Association	32
3	IBM Keypunch Card with Coded Data	33
4	Relationship of Regression Models I, II, and III	57

CHAPTER I

INTRODUCTION AND PURPOSE

Introduction

In recent years there has been an increasing awareness on the part of marketing practitioners and academicians that organizations outside of the traditional business setting could profit from the application of modern marketing tools and techniques to their operations. Professors' Kotler and Levy stimulated thinking in this area in an article which stressed the need for a broadening of the concept of marketing to include non-business organizations.[1] Even a superficial analysis of organizations like the Red Cross, Boy Scouts, and Heart Fund would indicate the complexity and breadth of their functions. Also, as the absolute number of these charitable organizations increases, the competition for the public's donation or assistance becomes increasingly intense. It becomes quickly evident that these groups do, in fact, have characteristics not dissimilar to their business counterparts and should be able to effectively utilize marketing methods that have proven successful in traditional profit-motivated organizations.

This study attempts to apply marketing analysis techniques to one such non-business organization, the Oklahoma Tuberculosis and Respiratory Disease Association, a chapter of the National Tuberculosis Association,

which is presently engaged in changing its name to the American Lung Association. This organizational name change, apparently without any formal market or image analysis, prompted the Oklahoma Chapter to request formal evaluation of their image, programs, and overall effectiveness. They were obliged to accept the national name change, but felt that at the state and local level their goals and objectives could be tailored to meet local demands. The Oklahoma Chapter approached the School of Journalism and Department of Administrative Sciences at Oklahoma State University in hopes of obtaining assistance in the form of applied field research.

What has evolved in response to this request is an interdepartmental study involving both academic disciplines. Specifically, the Department of Administrative Sciences study will involve three areas of market research, with a Doctoral candidate, two Master's candidates and a cooperating Professor participating in the project.

The Tuberculosis Association has relied exclusively upon a mail campaign for solicitation of donations, which includes the traditional "Christmas Seals" that accompany the solicitation letter. The seal and original slogan, "Stamp Out Tuberculosis," has been the central theme of their campaign for years. Now that the incidence of tuberculosis has been substantially reduced so that it no longer poses a threat to our country, the Association has directed much of its efforts towards other respiratory diseases and associated problems, such as emphyzema, lung research, clean air, and others. However, the Christmas Seal mail campaign remains as their major fund raising effort.

Due to the emphasis placed on mail solicitation by the Tuberculosis Association, the availability of socioeconomic and demographic data from the 1970 Census, and published data from the Internal Revenue Service, for each of the country's 38,000 5-digit zip code areas, it was concluded that a fertile area of research would be a statistical analysis of this data and the data from the records of the State Tuberculosis Association.

Objectives of the Study

The objective of this study was to identify the characteristics that are indicative of a successful zip code area, in terms of solicitation returns, so that a basis for prediction could be formed. In this manner, the results of the study could be used to evaluate the potential of a given zip code list of prospective donors. Also, the presently used mass solicitation techniques could be made selective so as to predict the most returns for campaign dollar spent, by defining "heavy giver" areas.

Following a review of pertinent literature, statement of the problem involved in this study, and discussion of the methodology to be employed, the data will be analyzed. The results of the analysis for practical use by the Tuberculosis Association, along with suggestions for future research in the area, will conclude this study.

FOOTNOTES

[1]Philip Kotler and Sidney J. Levy, "Broadening the Concept of Marketing," Journal of Marketing, (January 1969), p. 10.

CHAPTER II

STATEMENT OF THE PROBLEM

The central problem of this study was to determine whether or not various socioeconomic and demographic data available through the 1970 Census and the Internal Revenue Service could be used as accurate predictors of state zip codes that would have a high incidence and/or value of monetary returns from the Oklahoma Tuberculosis Association's mail solicitation campaign. The Oklahoma Chapter had indicated that they had a practical feeling for what has been in the past, at least, a "good" zip code in terms of solicitation response. There has not been, however, any formal analysis of their records to support this opinion.

With prediction then, as the prime objective to the study, it was felt that an analysis of dependence should accurately forecast key variables. Oklahoma State University is fortunate to have in its computing facility, a library of various multivariate statistical routines without which the study could not have been undertaken.

The Oklahoma Tuberculosis Association maintains past campaign records of its donors at offices in Oklahoma City and Tulsa. Oklahoma City handles 78 of the state's 87 counties, with the Tulsa Chapter maintaining the rest. Each office has a clerical staff which continually maintains the records system. The donor records of the Association are

separated into the 700+ 5-digit zip code areas within the State of Oklahoma. Each donor is indicated by a file card that has pertinent information, as to address, donation history, and amount donated. Specific reference will be made to this card file system later in the study.

Due to the availability of computer facilities and statistical programs, the form in which the O.T.B.R.D.A. kept its records, and the availability of data categorized by zip code from the 1970 Census and the Internal Revenue Service, it became apparent that the task of collecting and analyzing data would be within practical time and financial limits.

The problem of segmenting and identifying markets by zip code is significant from a number of aspects. First, those involved in profit-motivated organizations have already realized the benefits from zip code marketing. As is indicated in a recent publication by Time magazine, it can make available on a cost basis, detailed information of its circulation and distribution for all of the 38,000 United States zip code areas.[1] The article points out the profit potential and increase in efficiency that will result if a firm can determine those areas that appear to contain people most likely to purchase their product or service. Selective promotion techniques could then by applied to gain the greatest efficiency of the advertising dollar.

Second, the Oklahoma Tuberculosis Association relies almost entirely upon a mail solicitation campaign, so that after initially reviewing their methods of developing mailing lists and how they have evaluated

the efficiency of their mail campaign to date, the researchers felt that a detailed analysis might be able to suggest ways to lower costs and to improve the efficiency of the mail operation.

Third, very little work has been done in the area of selective promotion for charitable organizations, these groups almost totally relying on "mass" solicitation techniques. A study by Mindak and Bybee,[2] in which they applied marketing concepts to fund raising activities of the March of Dimes in a Texas county, suggested than an analysis that identified certain levels of donors by zip code, could be used to effectively reach "heavy giver" areas.

Fourth, due to the limited experience by marketing practitioners and researchers in the area of zip code analysis, this study to a large degree is of an introductory or pilot nature and should lay the ground work for further analysis.

In summary, this study opens the door to a relatively new area of marketing research, especially where it applies to charitable or non-profit motivated organizations. As the competition for charitable contributions increases, charitable groups will be by necessity, required to strive for the most efficient and economical way to locate and solicit contributions. With this study, the researchers hope to contribute in some small way to the overall improvement of the Oklahoma Tuberculosis and Respiratory Disease Association's continuing search for funds.

FOOTNOTES

[1]Time Inc., Profit from Zip Data, Time-Life Inc., (Boston, 1973).

[2]William Mindak and Malcolm Bybee, "Marketing's Application to Fund Raising," Journal of Marketing, Volume 35 (July 1971), p. 13.

CHAPTER III

SURVEY OF THE LITERATURE

Introduction

Two areas of literature review are relevant to the study. The first being marketing techniques as applied to non-profit organizations, of which there are very few published articles. Second is the use of Census and related statistical data in marketing research, a topic area in which articles abound.

Marketing's Application to Non-Business Organizations

Drs' Kotler and Levy were probably the first to point up the need to transfer traditional marketing principles to the functions of non-business organizations. The authors see a great opportunity for marketers to expand their thinking and to apply their skills to an increasingly interesting range of social activity.[1] They stress that the marketing tools which have been successfully applied to business firms have counterpart application to the non-business organizational activity. It will be to the benefit of these non-business organizations to accept and utilize marketing tools and techniques to more effeciently operate, for as Kotler and Levy point out, "No organization can avoid marketing, the choice is to do it well or poorly."[2]

No organization, whether profit-motivated or charitable, can function effectively without giving consideration to the application of marketing techniques. Business organizations have long realized the need for formalized marketing approaches. Charitable groups have for the most part, ignored placing any emphasis on an organized marketing activity, choosing rather to function with a hit or miss, shotgun type of appeal for funds.

Utilization of these techniques can not be totally unexpected. For many of these people their only approach to a formal concept of marketing was through exposure to books like the "Hidden Persuaders," and other Madison Avenue references to marketing as being a "social bad." It is no wonder that many of those involved in coordinating the activities of charitable organizations are skeptical of people who suggest the application of organizational marketing techniques to their operations. If these people can be convinced that their charity parallels the activities of a business of equal magnitude, that they have "products or services" to sell, that they have consumers or publics to deal with, and that they must be in constant communication with their environment, then as marketing has proven successful in business, it very well could prove to be successful in non-profit organizations. The task of broadening the use of marketing to non-business organizations lies with the students and practitioners of marketing. For as soon as the results of these organized marketing efforts begin to show up in the results of charitable campaigns, marketers will have little trouble in obtaining project groups.

Two marketing researchers recently took heed of Professors Kotler and Levy's appeal to apply marketing techniques to non-business organizations and conducted a study for the March of Dimes Foundation in a Texas county. Professor William Mindak and H. Malcolm Bybee of the University of Texas set out to answer a number of questions pertaining to the effectiveness of applying marketing to non-business enterprises. The field study and subsequent results were quite startling.

The particular study focused on a March of Dimes fund raising drive held in Travis county, Texas in the spring of 1970. This study was one of the first to apply marketing concepts to an area traditionally considered not to be a business enterprise.

The authors found the organization's handbook sorely outdated and the record keeping to be grossly inadequate and at times non-existent. They were able to develop, however, an indication of the problems confronting the March of Dimes, and through a preliminary market analysis a number of potential opportunities for improvement were revealed.

The analysis included application of a number of the techniques discussed by Kotler and Levy.[3] These included; a market segmentation on a heretofore undifferentiated campaign; a search for themes that accurately projected the March of Dimes current image; training sessions for those involved with the Mothers' March, in order to formalize the concepts of people solicitation techniques; and a marketing audit, to determine the effectiveness of the researchers' efforts.

The results of the study of Travis County were astonishing. Income (donations) increased 33% over the previous year and it was the first time in twelve years that contributions had increased. The implications of the study were the authors' conclusion that the charity should strive for a more clearly differentiated market and define the "heavy giver." They suggested a computer analysis of data from the Internal Revenue Service and Census Bureau in relation to the records of the charity as to donation size, etc.[4] This study will follow that line of thinking to a large extent. In the authors estimation they felt, "the results of the Travis county 'test market' clearly suggest that marketing techniques and philosophy can be applied to ideas and social causes. It also seems clear that other foundations such as the March of Dimes would profit through the application of such techniques."[5] The authors also suggest further study into selective promotion for charitable groups. They suggest the use of computer data from the Census Bureau and the Internal Revenue Service, complied by zip code areas.[6]

Use of Demographic and Socioeconomic Data in Marketing Research

Census Data

The following paragraphs will discuss a number of pertinent articles concerned with the opportunities for research afforded the marketing researcher by the availability and variety of socioeconomic and demographic data.

The decennial census conducted by the United States Government has long been recognized as a source containing a wealth of data. As early as 1954, N. H. Borden wrote in the Journal of Marketing about the possibilities of advancing marketing research through the use of quantitative techniques such as regression and correlation analysis in connection with census data, to test factors that might have an influence over the firm's sales. Borden did not, however, give any explicit information as to how the census data was used.[7] In 1958, the American Marketing Association discussed market potentials and the use of the census data in a journal report.[8]

Not until the 1960 Census however, did marketers realize the full potential and importance of the Census data to the furthering of marketing research efforts. Advances in computer technology during the 1960s had allowed more rapid tabulation and disemmination of Census data than had ever before been possible.

Another system that was to greatly assist data collection and measurement was the introduction of the United States Postal Zip Code System. Here the entire continental United States was broken down into 5-digit identification numbers, each signifying a certain density of population that would afford the Post Office with a particular amount of mail flow and that the area so signified contributed to the efficient transportation of the mails. The United States was ultimately divided into 38,000 5-digit zips which have proven invaluable to marketing research that concerns itself with any kind of mail solicitation or advertising, as was the case with this particular study.

In 1969, Dr. George H. Brown was picked to head the Bureau of the Census as it set out to undertake the task of developing the 1970 Census. Prior to coming to the Census Bureau, Dr. Brown had been the director of marketing research at Ford Motor Company. He had a high degree of empathy for the needs of marketers in the area of data collection and utilization, and he has vowed to make all the data that the Census has the potential of delivering, available to all those who desire it. To accomplish this the Bureau brought together a great wealth of people, computers, and financing. Over 8,000 Census Bureau workers, seemingly as many computers, and a budget of over $200 million dollars embarked upon the task of counting and categorizing the entire United States.[9] What has involved to present is the most sophisticated, complete, and most available collection of Census information ever obtained. A vast array of printed reports, summary computer tape files, data collection, and retrieval programs, mapping procedures, special request print-outs, and valuable Census Bureau field office assistance to users, are but a few of the services offered in regards to the 1970 Census. The advance of computer technology during the last 15 years has allowed the Bureau to make 3 to 4 times the data available as during a similar time span at the time of the 1960 Census.[10] One of the newest features of the 1970 Census was the sponsoring of over 125 Statistical User Service Centers, where computer services relating to data collection and retrieval are available. Oklahoma State University was fortunate to be chosen as one of these centers. Dr. Brown has pledged the Census Bureau to provide the most efficient and complete data system ever assembled, and has

carried his message to numerous business and related interest groups. A selected list of his publications appears in the bibliography.

United States Internal Revenue Service Tax Data

Another very important source of statistical data which is divided by 5-digit zip code is the Internal Revenue Service. In 1969 they undertook the herendous task of classifying every taxpayer for the 1969 tax year by the 5-digit zip code reported on his income tax form. Information was summarized as to income classes, type of return, number of exemptions and dependants, adjusted gross income, whether interest and/or dividends were claimed and the amount, and the amount of tax paid to the Federal Government. Each 5-digit zip code for the entire United States is summarized in this manner. This information is considered to be a valuable and reliable source of economic data, stemming from the viewpoint that most people report honestly on their tax returns. The data is contained in printed reports for each state developed for a master IRS File, and may be obtained from the source referenced in the Appendix.

As of this date the Internal Revenue Service indicated that it would be updating these reports for the 1972 tax year.

FOOTNOTES

[1]Philip Kotler and Sidney Levy, "Broadening the Concept of Marketing," Journal of Marketing, Volume 33 (January 1969), p. 10.

[2]Ibid., p. 15.

[3]Ibid.

[4]William Mindak and Malcolm Bybee, "Marketing's Application to Fund Raising," Journal of Marketing, Volume 35 (July 1971), p. 18.

[5]Ibid., p. 17.

[6]Ibid., p. 18.

[7]N. H. Borden and C. W. Smith, "An Appraisal of Census Programs for Marketing Uses," Journal of Marketing, (April 1954).

[8]Timothy R. Heyman, "Use of Census Data in Interregional Marketing," Sloan Management Review, Volume 12 (Winter 1971).

[9]Dr. George Brown, "Handier and Dandier Data for Marketers," Sales Management, (December 1970), p. 21.

[10]Ibid.

CHAPTER IV

RESEARCH METHODOLOGY

Determining the Sample

Oklahoma is divided in 756 5-digit zip codes in which the population for each zip varied from a high of 67,000 to a low of 19 and can be broken down as follows:

TABLE I

OKLAHOMA 5-DIGIT ZIP CODES
CATEGORIZED BY POPULATION LEVELS AND
SUMMARY TOTALS FOR EACH LEVEL

Population	Zip Areas Included
Over 10,000	61
Between 5 - 10,000	49
Between 1 - 5,000	213
Between 500 - 1,000	145
Under 500	288
	756 Total Zips

For this analysis it was decided to exclude zips in the "under 500 areas." This did eliminate some 288 zips, but it was concluded that

these small areas did not contribute a significant amount of information pertinent to the study. In reviewing the Association's records, it was found that these small areas had few, if any donors, and they could not have been used in the study for this reason.

A sample of 109 zips from the remaining 4 population segments was taken on the basis of 1 out of 3 in the "over 10,000" and "5 - 10,000" and 1 out of 5 in the "1 - 5,000" and "500 - 1,000" segments. In this way the disparity between numbers of zips in the population classes did not cause a bias in the choice of the sample.

The following is a list of zips selected and the corresponding city and county that identifies the zip area.

TABLE II

LIST OF 5-DIGIT ZIP CODES SELECTED FOR THE STUDY

ZIP	CITY	COUNTY
73501	Lawton	Commanche
73110	Midwest City	Oklahoma
73112	Thirty Ninth St., Oklahoma City	Oklahoma
74601	Ponca City	Kay
73107	Farley, Oklahoma City	Oklahoma
74801	Shawnee	Pottawatomie
73120	Village	Oklahoma
74114	Ranch Acres, Tulsa	Tulsa

TABLE II "Continued"

ZIP	CITY	COUNTY
73159	Will Rogers, Oklahoma City	Oklahoma
73109	Capitol Hill, Oklahoma City	Oklahoma
74820	Ada	Pontotoc
74066	Sapulpa	Creek
74102	Broken Arrow	Tulsa
74110	Northside, Tulsa	Tulsa
74063	Sand Springs	Tulsa
73108	Stockyards, Oklahoma City	Oklahoma
74017	Claremore	Rogers
74129	Southwest, Tulsa	Tulsa
73116	Lakeside, Oklahoma City	Oklahoma
74701	Durant	Bryan
74868	Seminole	Seminole
74023	Cushing	Payne
73601	Clinton	Custer
74437	Henryetta	Okmulgee
73020	Choctaw	Oklahoma
73132	Thirty Ninth St., Oklahoma City	Oklahoma
73130	Midwest City	Oklahoma
73139	Capitol Hill	Oklahoma
74467	Wagoner	Wagoner
74728	Broken Bow	McCurtain

TABLE II "Continued"

ZIP	CITY	COUNTY
74960	Stilwell	Adair
74021	Collinsville	Tulsa
74848	Holdenville	Hughes
74948	Muldrow	Sequoyah
73055	Harlow	Stephens
73104	Downtown Carriers, Oklahoma City	Oklahoma
74108	Admiral, Tulsa	Tulsa
74873	Tecumseh	Pottawatomie
73446	Madill	Marshall
74462	Stigler	Haskell
73010	Blanchard	McClain
73438	Healdton	Carter
74079	Stroud	Lincoln
73045	Harrah	Oklahoma
73550	Hollis	Harmon
74436	Haskell	Muskogee
74103	M. O. Carriers, Tulsa	Tulsa
74331	Afton	Ottawa
74441	Hulbert	Cherokee
74851	McLond	Pottawatomie
73089	Tuttle	Grady

TABLE II "Continued"

ZIP	CITY	COUNTY
74130	Northside, Tulsa	Tulsa
74002	Barnsdall	Osage
73763	Okeene	Blaine
74132	West Tulsa	Tulsa
73834	Buffalo	Harper
73860	Waynoka	Woods
74365	Salina	Mayes
73047	Hinton	Caddo
73669	Thomas	Custer
74343	Fairland	Ottawa
73527	Cache	Commanche
74561	Quinton	Pittsburg
74039	Kellyville	Creek
74880	Weleetka	Okfuskee
74733	Colbert	Bryan
74469	Warner	Muskogee
73724	Canton	Blaine
73628	Cheyenne	Roger Mills
74053	Oolagah	Rogers
74369	Welch	Craig
73565	Ryan	Jefferson
73843	Gage	Ellis

TABLE II "Continued"

ZIP	CITY	COUNTY
74451	Park Hill	Cherokee
74302	Glencoe	Payne
74857	Newalla	Oklahoma
74470	Webbers Falls	Muskogee
74867	Sasakwa	Seminole
74932	Cameron	Le Flore
74072	South Coffeyville	Nowata
74027	Delaware	Nowata
73560	Olustee	Jackson
74759	Soper	Choctaw
73027	Coyle	Logan
74734	Eagletown	McCurtain
73042	Gracemont	Coddo
74026	Davenport	Lincoln
73950	Turpin	Beaver
74735	Fort Towson	Choctaw
74824	Agra	Lincoln
74956	Shady Point	Le Flore
73016	Cashion	Kingfisher
73569	Terral	Jefferson
73625	Butler	Custer
73647	Foss	Washita
73450	Milburn	Johnston

TABLE II "Continued"

ZIP	CITY	COUNTY
73754	Lahoma	Garfield
74852	Macomb	Pottawatomie
74572	Tupelo	Coal
74428	Council Hill	Muskogee
73544	Gould	Harmon
74736	Garvin	McCurtain
74442	Indianola	Pittsburg
73661	Rosky	Washita
73660	Reydon	Roger Mills
74574	Tuskahoma	Pushmataha
74138	Admiral, Tulsa	Tulsa
73443	Lone Grove	Carter

Table III shows a list of the counties appearing in the sample and the number of times each appeared. It is interesting to note that 53 of the state's 87 counties are represented in the sample, with the majority of counties appearing no more than twice, except for the much larger (population-wise) Tulsa and Oklahoma counties. This sample can be considered as a valid cross-section of the state.

TABLE III

ALPHABETICAL LIST OF COUNTIES IN SAMPLE AND NUMBER OF TIMES APPEARED

	County	Number of Times Appeared
1.	Adair	1
2.	Beaver	1
3.	Blaine	2
4.	Bryan	2
5.	Caddo	2
6.	Carter	2
7.	Cherokee	1
8.	Choctaw	2
9.	Coal	1
10.	Commanche	2
11.	Craig	1
12.	Creek	2
13.	Custer	2
14.	Ellis	1
15.	Garfield	1
16.	Grady	1
17.	Harmon	2
18.	Harper	1
19.	Haskell	1

TABLE III "Continued"

	County	Number of Times Appeared
20.	Hughes	1
21.	Jackson	1
22.	Jefferson	2
23.	Johnston	1
24.	Kay	1
25.	Kingfisher	1
26.	Leflore	2
27.	Lincoln	3
28.	Logan	1
29.	Marshall	1
30.	Mayes	1
31.	McClain	2
32.	McCurtain	3
33.	Muskogee	5
34.	Nowata	2
35.	Okfuskee	1
36.	Oklahoma	15
37.	Okmulgee	1
38.	Osage	2
39.	Ottawa	2
40.	Payne	2
41.	Pittsburg	2

TABLE III "Continued"

	County	Number of Times Appeared
42.	Pontotoc	1
43.	Pottawatomie	4
44.	Pushmataha	1
45.	Roger Mills	2
46.	Rogers	2
47.	Seminole	2
48.	Sequoyah	1
49.	Stephens	1
50.	Tulsa	11
51.	Wagoner	1
52.	Washita	2
53.	Woods	1

Collection of Data

After selection of the zip codes to be included in the sample, the data was collected and coded for computer analysis.

Independent Variable

It was originally proposed to use summary data from the "Fifth Count Tape" of the 1970 Census as a basis from which to develop our independent variables. This tape contains social and economic data

classified by the 5-digit zip codes of the United States Postal System. However, at the time this study was being developed, the researchers did not have access to this information. The computing facility is scheduled to receive the Fifth Count Summary tapes for Oklahoma and surrounding areas as soon as they become available.

The alternate source of data for the independant variables was a printed report of the Internal Revenue Service's Individual Income Tax Data for Each 5-Digit Zip Code Area in Oklahoma - Tax Year 1969. The data is purely economic in nature, but it was felt that a high level of predictability concerning donations could be obtained through the use of this information.

The information contained in this report, although limited to that which is contained on an individual's Federal Income tax return, would tend to be a very significant source of data for the study. This could be justified by the assumption that donations are highly correlated to income and related personal statistics. The report has a high measure of reliability and validity due to the penalities associated with submitting a fraudulent tax return.

Each 5-digit zip code contains summary data for all individual income tax returns that had indicated as coming from that zip. If a return did not contain a zip. the I.R.S. computer was able to determine from the address what the proper zip code was and affix it to the return. Through this procedure it was possible to place almost 100% of the tax returns into a five digit zip code.

The I.R.S. report classified data as follows. First, it divided income as reported on the return into five classes:

Under $3,000
Between $3,000 and $5,000
Between $5,000 and $10,000
Between $10,000 and $15,000
Over $15,000

It then categorized each one of the income levels as follows:

1. Number of Returns - This is the total number of individual tax returns filed.
2. Number of Joint Returns - This is the number of joint (husband-wife) tax returns filed.
3. Number of Exemptions - As claimed on the tax return - This figure was subdivided into taxpayer (personal, including blind, over 65, etc.) and dependent exemptions. This includes children and others who the taxpayer supports.
4. Adjusted Gross Income - The dollar amount reported by the taxpayer on line 18 of form 1040, and includes salary, wages, tips, and all other sources of income; less any expenses and exclusions to that income.
5. Dividends in Adjusted Gross Income - The number of returns claiming dividends and the total dollar amount of these dividends.
6. Interest Received - The number of returns claiming interest and the total dollar amount of interest received.

7. Total Tax - The Dollar amount of Tax paid to the Federal Government.

After determining amounts for each of the income levels, a total for each category was derived so that sum totals for the entire zip could be obtained.

A sample page for that report is shown in Figure 1.

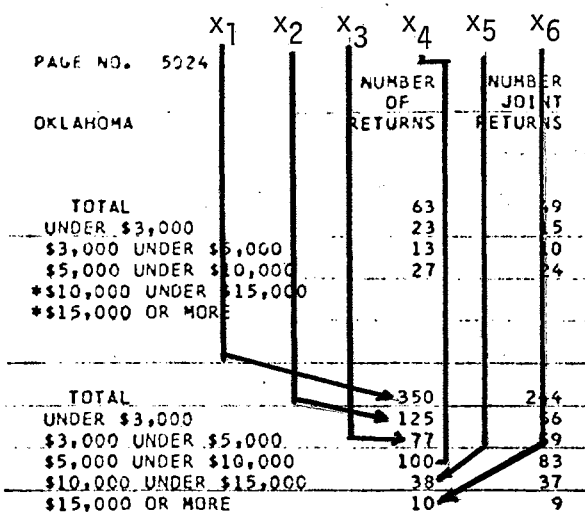
It was a relatively simple matter to obtain data from this report and code it for the computer analysis, as the zip codes are in ascending numerical order.

The wealth of information in this report that is available to marketers cannot be overstated. Herein lies a number of the critical variables that indicate relative purchasing power of representative populations.

The Dependent Variables

The information for the dependent variables was obtained from the files of the Oklahoma Tuberculosis Association's Tulsa and Oklahoma City offices.

The donor records are contained in two file systems referred to as the "General File" and the "Special File." The "General File" contains a list of all donors and the "Special File" is a list of donors who gave $10 or more. Each donor is represented by a 3 x 5 file card. The general donor ($5 or less) was included in the general file on a white card with all pertinent information. A special donor was



PAGE NO. 5024

OKLAHOMA	NUMBER OF RETURNS	NUMBER JOINT RETURNS	NUMBER OF EXEMPTIONS TOTAL	TAXPAYER	DEPENDENT	ADJUSTED GR. INCOME (THOUSAND DOLLARS)	DIVIDENDS IN AGI NUMBER OF RETURNS	DIVIDENDS IN AGI AMOUNT (THOUSAND DOLLARS)	INTEREST RECEIVED NUMBER OF RETURNS	INTEREST RECEIVED AMOUNT (THOUSAND	TOTAL TAX (DOLLARS)
73001 ALBERT						CADDO					
TOTAL	63	49	181	128	53	297	3		13	5	31
UNDER $3,000	23	15	52	50	2	31			6	3	1
$3,000 UNDER $5,000	13	10	33	25	8	51			2		4
$5,000 UNDER $10,000	27	24	96	53	43	215	3		5	2	26
*$10,000 UNDER $15,000											
*$15,000 OR MORE											
73002 ALEX						GRADY					
TOTAL	350	244	993	672	321	1830	8	9	103	90	205
UNDER $3,000	125	56	266	219	47	178	4	1	33	17	8
$3,000 UNDER $5,000	77	59	232	157	75	306			18	13	21
$5,000 UNDER $10,000	100	83	325	198	127	721	2	1	31	43	82
$10,000 UNDER $15,000	38	37	135	76	59	453	2	7	15	10	63
$15,000 OR MORE	10	9	35	22	13	172			6	7	31
73003 ALMA						STEPHENS					
TOTAL	270	187	772	489	283	1581	4	1	68	32	224
UNDER $3,000	105	45	205	172	33	151	1	1	20	6	12
$3,000 UNDER $5,000	32	21	102	56	46	129	1		5	4	8
$5,000 UNDER $10,000	104	94	361	200	161	774	1		31	6	80
$10,000 UNDER $15,000	29	27	104	61	43	527	1		12	16	125
*$15,000 OR MORE											
73004 AMBER						GRADY					
TOTAL	175	125	538	322	216	913	3	2	44	23	97
UNDER $3,000	54	25	116	92	24	70			9	4	3
$3,000 UNDER $5,000	42	29	122	77	45	167			13	5	13
$5,000 UNDER $10,000	58	51	221	109	112	417	2	2	10	7	45
$10,000 UNDER $15,000	21	20	79	44	35	259	1		12	7	36
*$15,000 OR MORE											
73005 ANADARKO						CADDO					
TOTAL	2846	1730	7941	5023	2918	16313	121	101	772	530	1927
UNDER $3,000	1080	321	2025	1618	407	1443	16	3	192	86	55
$3,000 UNDER $5,000	500	320	1467	909	558	2018	23	16	112	76	144
$5,000 UNDER $10,000	831	685	2894	1605	1289	5894	26	16	226	117	624
$10,000 UNDER $15,000	296	272	1074	593	481	3565	19	10	133	79	494
$15,000 OR MORE	139	132	481	298	183	3393	37	56	109	172	632
73006 APACHE						CADDO					
TOTAL	937	640	2770	1730	1040	5144	24	9	204	169	582
UNDER $3,000	308	120	630	508	122	456	7	2	62	33	27
$3,000 UNDER $5,000	177	119	510	330	180	712	2		44	41	57
$5,000 UNDER $10,000	343	296	1220	664	556	2471	6	4	55	31	264
$10,000 UNDER $15,000	86	82	330	176	154	1003	6	2	25	30	142
$15,000 OR MORE	23	23	80	52	28	502	3	1	18	25	101

was included in the general file on an orange card by name and address only and had to be cross referenced to a special file for the history of donations. Donors in the "Special File" were given preference as to mailing and handling, solicitation techniques, and materials sent during campaigns. Both files contained identical information pertaining to the history of donations.

The records in the General File are broken down by county, which is further subdivided into numerical zip codes within a county. For the majority of counties, a zip would represent a city, i.e., Payne - 74074 - Stillwater. For the larger cities (Tulsa and Oklahoma City), a zip code indicates a section of that city, i.e., Oklahoma City - 73108 - Stockyards. Each zip contains an alphabetical list of donors with a file card for each donor. The card shows name and address, history of donations (year and dollar amount), and a code relative to the amount of donations and whether or not the account is current. The Special File was organized alphabetically by county, and within each county the donors were listed alphabetically. This file was not broken down on the basis of zip code and made collection of data on this group difficult. Information contained on the cards in the "Special File" was identical to that in the "General File."

Figure 2 is a sample file card. The yearly donation is tabulated manually by the clerical staff each spring when they post the returns for the Christmas Seal campaign. The code (C-3) is an indication of the donation, and varies from donor to donor.

KEY C 3

Mrs Ima Donor
1212 Charity Lane
Contribution, Oklahoma 73193

YEAR	AMT.	Remarks	YEAR	AMT.	Remarks	YEAR	AMT.	Remarks	YEAR	AMT.	Remarks
69	1.00										
70	1.00										
71	2.00										
72	3.00										
73	3.00										

Figure 2. Donor File Card

It was decided to "pull" a systematic sample of 30 donors from each zip in the sample and record the total donations for those 30 donors. Total donors in a zip were estimated by measuring the length of the file index for that zip by considering 100 file cards equal to one (1) inch.

Because the study was concerned primarily with individual donors, i.e., households, file cards that referenced a business or firm were passed over in the gathering of data from chapter records. In other words, if in the sequence of the draw, a card was chosen that referenced a business donation, that card was passed over and the next card chosen. This procedure did not decrease the validity of the results because the data for the independent variables was based on individual tax returns only.

Another qualification of the data collection was that the special file, i.e., donors giving $10 or more, was not included in the analysis.

This was primarily due to the manner in which that group was filed. To cross reference a special donor drawn in the sample of the general file would have been an exceedingly difficult and time consuming task. The deletion of this segment of the donor population could be argued as possibly limiting the validity of the analysis. However, the special file comprised only 2% of the total donor population and the study of records of the Association indicated that where a high incidence of special donors was found, a correspondingly high incidence of $5 donations was also found, which were included in the study. This finding confirmed the fact that the method of data collection accurately described the sample of zip code areas.

After the information had been gathered from the I.R.S. Report and the Tuberculosis Association's records, it was keypunched onto I.B.M. data cards for ease of computer analysis. Figure 3 shows a sample data card with the sequence of information groupings indicated.

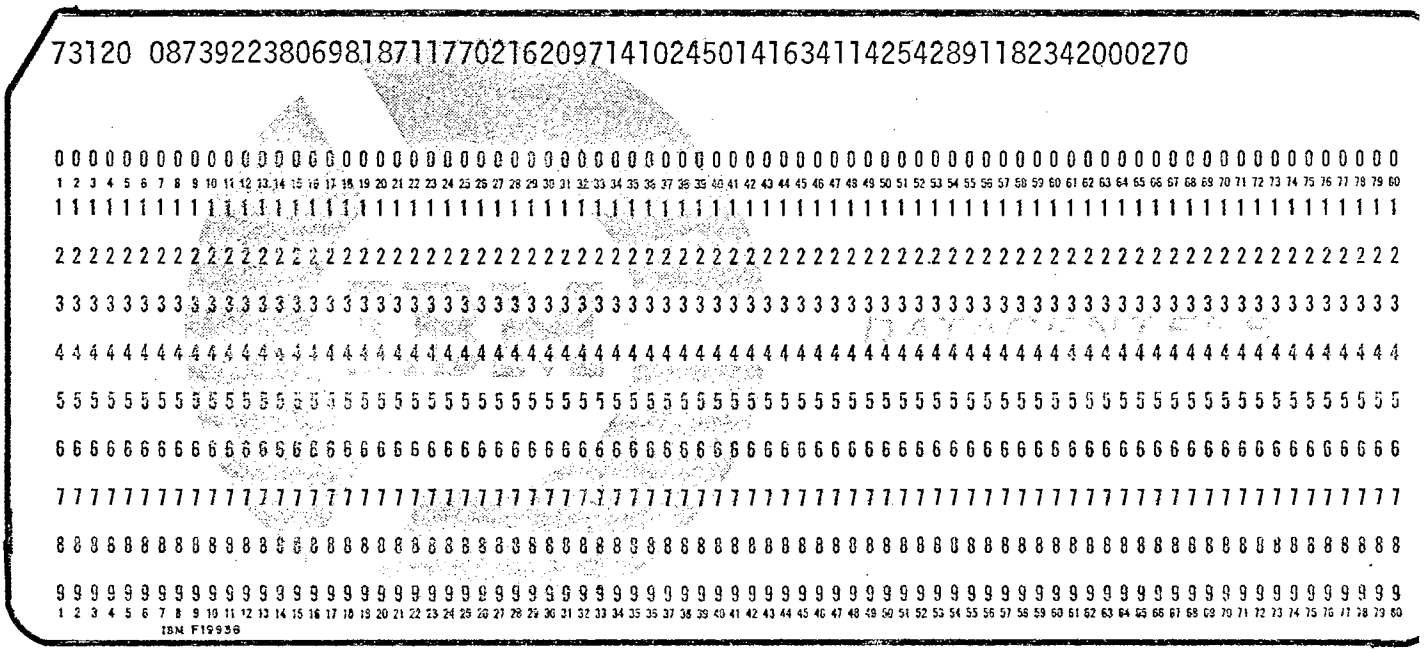


Figure 3. IBM Keypunch Card with Coded Data

Method of Analysis

The objective of the study was to identify profitable zip codes and attempt to predict other zips, because of similiar characteristics, that appear to have the potential of being "heavy giver" areas. The data was in the form of a number of independent variables based on the Internal Revenue Service's Report, and the dependent variable(s) which were our sample information from the Tuberculosis Association.

For analyzing dependence, regression analysis is the most commonly used technique. Its underlying theory is also the most developed. In regression analysis, a single, interval scaled dependent variable is to be predicted or explained by a set of independent variables which are assumed to be interval scaled.[1]

For this analysis a stepwise regression analysis was chosen. The following paragraphs will briefly describe this regression program. The reader is directed to Dixon,[2] for a complete explanation of this and other computer programs.

BMD02R - STEPWISE REGRESSION

General Description

1. This program computes a sequence of multiple linear regression equations in a stepwise manner. At each step one variable is added to the regression equation. The variable added is the one which makes the greatest reduction in the error sum of squares. Equivalently it is the variable which has highest partial correlation with the dependent variable partialed on

the variables which have already been added; and equivalently it is the variable which, if it were added, would have the highest F value. In addition, variables can be forced into the regression equation. Non-forced variables are automatically removed when their F values become too low. Regression equations with or without the regression intercept may be selected.

2. Output from this program includes:

   A. At each step:

      (1) Multiple R

      (2) Standard error of estimate

      (3) Analysis-of-variance table

      (4) For variables in the equation:

         (a) Regression coefficient
         (b) Standard error
         (c) F to remove

      (5) For variables not in the equation:

         (a) Tolerance
         (b) Partial correlation coefficient
         (c) F to enter

   B. Optional output prior to performing regression:

      (6) Means and standard deviation

      (7) Covariance matrix

      (8) Correlation matrix

   C. Optional output after performing regression:

      (9) List of residuals

      (10) Plots of residuals vs. input variables

      (11) Summary table

This program was run on Oklahoma State University's IBM Model 360-65 computer. Checking the compatibility of this program to other models and other manufacturer's hardware is suggested and can be found in Dixon.[3]

The stepwise regression analysis was used to hopefully "discover" the most important variables. However, it must be pointed out that the variables are "discovered" because they appear to be the best predictors of a "heavy giver" zip code areas, and not because there is necessarily any causality between them and heavy givers.

FOOTNOTES

[1]David A. Aaker, Multivariate Analysis in Marketing: Theory and Practice, (1971), p. 3.

[2]W. J. Dixon, BMD - Biomedical Computer Programs, (1971), p. 233 - 250.

[3]Ibid.

CHAPTER V

ANALYSIS OF DATA AND INTERPRETATION OF RESULTS

This chapter will first discuss the variables used in the analysis, then describe the results obtained in the various computer runs with the development of three prediction models. Finally, an overall discussion of the models in terms of practical application will conclude this chapter.

Independent Variables

The independent variables as previously mentioned, were developed exclusively from the Internal Revenue Service's Tax Report and reflect statistics on income and related factors from the 1969 tax year. It was felt that information of the type that is contained in a person's income tax return would tend to be a good predictor of the propensity for a person to donate to charitable organizations, as many of these facts have previously been used to predict areas which might have a high potential of purchasing power for the benefit of profit-motivated organizations. The level of income is highly related to the overall ability to purchase or donate, and the supplemental information on dividends and interest considered to be other sources of income, might point to an individual who has excess funds that might be diverted

into a charity. Information regarding exemptions, both personal and dependent, would indicate whether those families with or without children might be more inclined to donate.

The independent variables were broken into two groups or categories. Group 1 being a raw data as was obtained from the tax report. Each variable is indicative of an absolute dollar amount or number of returns (families) in that category. It should be pointed out that, generally speaking, a return refers to a family unit, as during 1969 a joint return would have usually given a more favorable tax rate for the family than the husband and wife filing separately.

Group 2 is a transformation of the original variables through the regression program and represent percentage or averages of the total families or returns as indicated for a zip area.

Table IV lists and explains the two groups of independent variables as they were coded and defined in the regression analysis.

TABLE IV

EXPLANATION OF VARIABLES

Code		Definition
		Group 1
1.	TOTFAM	Total individual tax returns filed (tax year 1969) for given zip
2.	UNDER 3	Number of returns with income less than $3,000

TABLE IV "Continued"

	Code	Definition
3.	BTWN 35	Number of returns with income between $3,000 and $5,000
4.	BTN510	Number of returns with income between $5,000 and $10,000
5.	BN1015	Number of returns with income between $10,000 and $15,000
6.	GRTR15	Number of returns with income greater than $15,000
7.	DEPEXP	Number of dependent exemptions claimed on the total returns for given zip
8.	AGI	Total adjusted gross income for a given zip (in 000's of dollars)
9.	DIVFAM	Number of returns claiming dividends in given zip
10.	DIVDOL	Total dollar amount of dividends claimed in given zip (000's of dollars)
11.	INTFAM	Number of returns claiming interest received for given zip
12.	INTDOL	Total dollar amount of interest received in given zip (000's of dollars)
13.	TOTTAX	Total tax paid to government by all returns in given zip (000's of dollars)
		Group 2
16.	PCTUN3	Percent of returns with income under $3,000 for given zip (2/1)
17.	PCNT35	Percent of returns with income between $3,000 and $5,000 (3/1)
18.	PCT510	Percent of returns with income between $5,000 and $10,000 (4/1)

TABLE IV "Continued"

	Code	Definition
19.	PT1015	Percent of returns with income between $10,000 and $15,000 (5/1)
20.	PTOV15	Percent of returns with income over $15,000 (6/1)
21.	AVGAGI	Average adjusted gross income per family (return) (000's of dollars) (8/1)
22.	PCTDIV	Percent of families in given zip claiming dividends (9/1)
23.	AVGDIV	Average dollar amount of dividends received per family in given zip (000's of dollars) (10/1)
24.	PCTINT	Percent of families in given zip receiving or claiming interest (11/1)
25.	AVGINT	Average dollar amount of interest received per family in given zip (000's of dollars) (12/1)
26.	AVGTAX	Average tax paid per family (return) in given zip (000's of dollars) (13/1)
27.	AVGDEP	Average number of dependents exemptions claimed per family in given zip (7/1)

Dependent Variables

The role of the dependent variable was to be a fair representation of a "heavy giver" zip code. A "heavy giver" zip code could be defined as either having many donors giving various amounts of money or a small number of donors who gave large amounts of money.

The dependent variables were of two types; that which was an indication of people (donors) within a zip area and that which was an indication of the amount of money (donation) which can be attributed to that zip. Each of these could be determined from the Tuberculosis Association's records for each zip code in our sample population. It was felt that due to the population distribution, in order to accurately define and predict a "heavy giver" zip area, the study would have to include a combination of both donor (people) and donation (dollar) variables.

The original (raw) data came from the chapter records and was an indication of the number of donors and the amount they contributed per zip. Through the computer analysis it was possible to convert this data into terms relative to the population base of that zip. With this transformation routine, the raw data was transferred into averages and percentages, which were more meaningful than the raw data on a comparative basis. Because Oklahoma has a wide population disparity, it was felt that the percentage of families in a zip that donated would be extremely helpful, especially in the lesser populated areas.

Table V lists and defines the dependent variables as they were identified in the analysis. Group 1 contains the original data as obtained from the chapter records. Group 2 is the transformation of the original data by the computer program. The variables will maintain their same indices throughout the various regression models.

TABLE V

EXPLANATION OF THE DEPENDENT VARIABLES

Code		Definition
		Group 1 - Original Data in Raw Form
14.	DONORS	Number of donors in given sample zip
15.	DONAVG	Average dollar donation per donor for sample in given zip
		Group 2 - Transformed Data
28.	TOTDON	Total dollar amount of donations in a given sample zip in dollars (DONORS (14) times DONAVG (15))
29.	DONFAM	Average donation per family in given sample zip in dollars (DONORS (14) times DONAVG (15) divided by TOTFAM (1))
30.	PCTDON	Percent of families in given sample zip that are donors (DONORS (14) divided by TOTFAM (1))

Results of the Regression Analysis

As was stated earlier, the explanation and prediction of donations for the Oklahoma Tuberculosis Association would be considered from the standpoint of donors (people) and/or average donation per donor. This was primarily due to the population desparity within the State of Oklahoma.

The results of the regression analysis will be explained by developing three regression models. Each of the models will be defined by a different dependent variable. Model I will consider an explanation of the percentage of donors in a zip code, Model II will consider what variables explain the level of average donation per donor in a zip, and Model III will be developed from the standpoint of explanation and prediction of total donations from a zip code area. It was felt that with the use of three predictive models, a more accurate "feel" or understanding of the data analysis could be obtained.

Table VI presents a list of the variables used throughout the analysis and their corresponding indices. Also included are the means and standard deviation for each of the variables.

TABLE VI

INTRODUCTORY STATISTICS

STEPWISE REGRESSION ANALYSIS

Variable		Mean	Standard Deviation
TOTFAM	1	2342.92773	3319.31592
UNDER3	2	702.68018	917.57495
BTWN35	3	327.50513	431.22876
BTN510	4	718.86597	1012.65601
BN1015	5	383.84521	641.03735
GRTR15	6	210.39174	471.14746
DEPEXP	7	2288.57715	3071.64062
AGI	8	17640.76953	28270.91016
DIVFAM	9	166.67009	375.44507
DIVDOL	10	382.36060	1327.35083

TABLE VI "Continued"

Variable	Mean	Standard Deviation
INTFAM 11	805.77319	1372.36157
INTDOL 12	535.76270	1063.99292
TOTTAX 13	2539.49463	4544.81250
DONORS 14	479.02051	777.09448
DONAVG 15	2.38638	0.41711
PCTUN3 16	0.35055	0.08217
PCNT35 17	0.16828	0.04656
PCT510 18	0.31102	0.05876
PT1015 19	0.12260	0.06430
PTOV15 20	0.04791	0.06379
AVGAGI 21	6.29929	3.35556
PCTDIV 22	0.04585	0.05501
AVGDIV 23	0.11889	0.62395
PCTINT 24	0.27836	0.10205
AVGINT 25	0.19405	0.31795
AVGTAX 26	0.84782	1.21232
AVGDEP 27	1.03310	0.17060
TOTDON 28	1220.94189	2160.56323
DONFAM 29	0.45320	0.18682
PCTDON 30	0.18859	0.06315

Table VII presents the entire correlation matrix for the analysis. The variables are indicated here by number only so that the reader is referenced to the preceeding table for their corresponding identification, and tables IV and V for a detailed explanation of each.

TABLE VII

CORRELATION MATRIX

Variable Number	1	2	3	4	5	6	7	8	9	10
1	1.000	0.933	0.956	0.984	0.959	0.833	0.973	0.962	0.792	0.404
2		1.000	0.983	0.975	0.897	0.764	0.941	0.916	0.759	0.368
3			1.000	0.970	0.845	0.671	0.901	0.858	0.682	0.296
4				1.000	0.930	0.733	0.965	0.909	0.689	0.304
5					1.000	0.878	0.966	0.963	0.779	0.408
6						1.000	0.812	0.943	0.935	0.647
7							1.000	0.940	0.714	0.351
8								1.000	0.894	0.584
9									1.000	0.771
10										1.000

Variable Number	11	12	13	14	15	16	17	18	19	20
1	0.965	0.787	0.874	0.935	0.195	-0.439	-0.437	-0.051	0.457	0.468
2	0.935	0.754	0.824	0.932	0.154	-0.345	-0.352	-0.059	0.354	0.397
3	0.884	0.705	0.753	0.885	0.102	-0.326	-0.269	0.003	0.300	0.308
4	0.913	0.703	0.796	0.896	0.138	-0.445	-0.408	0.053	0.446	0.369
5	0.945	0.756	0.880	0.888	0.237	-0.532	-0.549	-0.067	0.601	0.541
6	0.920	0.870	0.951	0.826	0.363	-0.446	-0.521	-0.269	0.483	0.714
7	0.913	0.703	0.842	0.884	0.205	-0.490	-0.494	-0.030	0.530	0.484
8	0.982	0.882	0.968	0.920	0.298	-0.488	-0.519	-0.156	0.505	0.640
9	0.913	0.961	0.936	0.861	0.351	-0.358	-0.423	-0.268	0.352	0.660
10	0.545	0.848	0.754	0.492	0.407	-0.269	-0.347	-0.352	0.193	0.728
11	1.000	0.887	0.936	0.960	0.271	-0.436	-0.460	-0.137	0.455	0.563
12		1.000	0.948	0.843	0.355	-0.368	-0.415	-0.266	0.325	0.692
13			1.000	0.864	0.370	-0.467	-0.521	-0.250	0.455	0.751
14				1.000	0.243	-0.373	-0.396	-0.115	0.389	0.480
15					1.000	-0.345	-0.342	-0.160	0.306	0.525
16						1.000	0.766	-0.445	-0.869	-0.564
17							1.000	-0.222	-0.862	-0.648
18								1.000	0.206	-0.391
19									1.000	0.558
20										1.000

TABLE VII "Continued"

Variable Number	21	22	23	24	25	26	27	28	29	
1	0.261	0.328	0.051	0.458	0.078	0.139	-0.235	0.904	0.259	0
2	0.208	0.306	0.038	0.412	0.068	0.111	-0.297	0.891	0.265	0
3	0.161	0.243	0.016	0.353	0.046	0.078	-0.318	0.835	0.214	0
4	0.209	0.236	0.010	0.387	0.034	0.094	-0.212	0.854	0.189	0
5	0.307	0.322	0.044	0.489	0.065	0.153	-0.115	0.872	0.240	0
6	0.423	0.546	0.187	0.605	0.211	0.281	-0.175	0.846	0.381	0
7	0.269	0.280	0.029	0.422	0.042	0.130	-0.101	0.857	0.213	0
8	0.422	0.494	0.196	0.572	0.220	0.290	-0.222	0.911	0.341	0
9	0.416	0.634	0.252	0.636	0.288	0.309	-0.322	0.877	0.478	0
10	0.707	0.839	0.680	0.617	0.682	0.678	-0.303	0.532	0.502	0
11	0.322	0.449	0.110	0.567	0.148	0.196	-0.278	0.949	0.372	0
12	0.567	0.723	0.449	0.661	0.482	0.492	-0.383	0.851	0.496	0
13	0.582	0.663	0.392	0.650	0.412	0.471	-0.267	0.869	0.424	0
14	0.267	0.412	0.086	0.530	0.128	0.159	-0.300	0.986	0.457	0
15	0.472	0.499	0.353	0.443	0.363	0.401	0.038	0.339	0.551	0
16	-0.496	-0.302	-0.143	-0.536	-0.151	-0.286	-0.294	-0.367	-0.115	0
17	-0.523	-0.356	-0.185	-0.517	-0.177	-0.325	-0.240	-0.402	-0.131	0
18	-0.221	-0.395	-0.311	-0.180	-0.319	-0.285	0.345	-0.148	-0.355	-0
19	0.392	0.194	0.006	0.473	0.005	0.161	0.333	0.384	0.070	-0
20	0.830	0.814	0.599	0.756	0.610	0.705	-0.093	0.514	0.494	0
21	1.000	0.824	0.907	0.622	0.900	0.966	-0.110	0.291	0.405	0
22		1.000	0.777	0.782	0.823	0.806	-0.425	0.449	0.641	0
23			1.000	0.423	0.984	0.982	-0.248	0.113	0.375	0
24				1.000	0.521	0.515	-0.384	0.548	0.699	0
25					1.000	0.973	-0.344	0.152	0.457	0
26						1.000	-0.210	0.182	0.395	0
27							1.000	-0.283	-0.477	-0
28								1.000	0.510	0
29									1.000	0
30										1

Table VII contains a high degree of intercorrelation between the variables, thus causing many of the signs and magnitude of the correlation to be erroneous. It was a repeated finding throughout the analysis that when a particular variable was allowed to enter, the next variable to enter would be the same in each case.

Econometric Models

Model I: Percent of Donors

$$Y = a_0 + a_1 x_1 + a_2 x_2 \ldots a_n x_n + E$$

where

Y = percent of families in zip that donated

$x_1 .. x_n$ = independent variables

$a_1 .. a_n$ = coefficients

E = error term

Model I attempted to explain the percent of families (donors) in a zip code. This relationship considered both donor and non-donor family units. Variables were allowed to enter naturally, (i.e., no force variables, and the level of significance for inclusion in the analysis was t = .05.)

Other independant variables that tend to correlate with the dependant variable but did not enter the analysis were as follows:

DIVFAM (9) - .340	TOTTAX (13) - .264
DIVDOL (10) - .317	PCTDIV (22) - .451
INTFAM (11) - .268	AVGINT (25) - .304
INTDOL (12) - .348	

NOTE: For an explanation of the above variables, see Table IV, p. 39-42.

The summary statistics for Model I are shown in Table VIII.

TABLE VIII

SUMMARY STATISTICS - MODEL I

VARIABLE **	MR	RSQ	INCRSQ	BETA COEFFICIENT	STD ERROR	B/B_S	F RATIO
Constant				-.096			
PCTINT (24)	.5889	.3468	.3468	.6315	.0669	9.44*	
PCTUN3 (16)	.7273	.5290	.1823	.3453	.0658	5.25*	38.6
PTOV15 (20)	.7444	.5541	.0250	-.2499	.1093	-2.29*	P .0001

* significance at the .05 level
** variables appear in the order in which they entered the analysis using a step-wise regression

"Percent Interest (24)" entered the equation first and would appear to indicate that families having income from sources other than salary or wages tend to be donors to the organization. Naturally, total income for a zip code is closely related to "percent interest" with high income zip code areas having high percentages. However, the entrance of interest into the model illustrates jointly a social class influence and an age of donor influence.

"Percent of Families under $3,000 (16)" (in income) entered the equation second. On the surface this relationship seems strange since it suggests that people in poverty levels are "heavy givers." However, the result must be interpreted in consideration of the fact that the interest variable had previously entered the model. By itself, variable 16 does not relate to the dependent variable. However, in this model

it appears likely that the "Under $3,000" refers to people in a fixed income situation such as older people living on interest from bank accounts, savings, and loans, etc.

"Percent of Families with income over $15,000 (20)" came in on the next step as is logical. However, "Over $15,000" has a negative coefficient indicating an inverse relationship to the dependent variable, (i.e., as the % of families with incomes over $15,000 increases, the % of families that donate decreases.) Independent of the other variables, high income has a positive relationship with donations. However, since it also interrelates with "Percent Interest," a negative sign best adds to the predictive power of the model.

In summary, the zip code with a high percent of donor families will be characterized primarily by

- a high percent of families with interest income.
- with a high percent of fixed income families (retired).
- or a high percent of high income families.

Therefore, based on the foregoing analysis, Model I can be stated in the following manner:

Model I

$$Y_1 = -.096 + .6315\, x_{24} + .3453\, x_{16} - .2499\, x_{20} + E$$

where

Y_1 = Percent of families in zip that donated

x_{24} = Percent of families in zip claiming interest

x_{16} = Percent of families in under $3,000 income bracket

x_{20} = Percent of families in over $15,000 income bracket

Model II: Donation per Donor

Model II attempted to identify those variables that are significant in pinpointing zip codes that will yield a high average donation per donor family. In this model, only donor populations were given consideration in the analysis. The results are not as definitive as other models since average donations could be identical for populations of high and low donor levels or for populations of uniform moderate levels. Only 27.6 percent of the variability in donation levels is explained. Variables were again allowed to enter naturally and the level of significance was still (t = .05.)

Other independent variables that tend to correlate with the dependent variable but did not enter the analysis were as follows:

GRTR15 (6) - .363	AVGAGI (21) - .472
DIVFAM (9) - .351	PCTDIV (22) - .499
DIVDOL (10) - .407	AVGINT (23) - .353
INTDOL (12) - .355	PCTINT (24) - .443
TOTTAX (13) - .370	AVGINT (25) - .363
	AVGTAX (26) - .401

NOTE: For an explanation of the above variables see Table IV, p. 39-42.

The summary statistics for Model II are shown in Table IX.

TABLE IX

SUMMARY STATISTICS - MODEL II

Variable**	MR	RSQ	INCRSQ	BETA COEFFICIENT	STD ERROR	B/B_S	F RATIO
Constant				2.228			
PTOV15 (20)	.5253	.2759	.2759	3.434	.5708	6.01*	36.196 P<.0001

* significance at the .05 level
** variables appear in the order in which they entered the analysis using a step-wise regression.

"Percent over $15,000 (18)" was the first and only variable to enter the equation that was statistically significant. However, this type of result would be expected in that those families with higher incomes that chose to donate would tend to donate more based on affluence, tax situation, and a host of other factors not considered in the analysis. Other variables that closely reflect the same result are "average adjusted gross income (21)" and "percent dividend (22)." Here we have the group of high, active wage and salary earners that make up a small percentage of the total donor population but are significant from the standpoint of the total dollars they contribute.

In summary, Model II simply states that the more the family income the higher the amount of donations per donor. This has special implications in seeking to enlarge the "special donor" mailing lists.

It must be stated at this point that this model is limited in its predictive capability because the data used came only from the

Tuberculosis Association's General File, i.e. contributions, under $10.00 had data from the Special file, contributions over $10.00, been utilized, this analysis would probably have shown better results.

Model II can be shown as follows:

$Y_2 = 2.228 + 3.434\ x_{20} + E$

where

Y_2 = Average dollar donation per donor

x_{20} = Percent of families in sample zip code with incomes over $15,000

E - Error term

Model III: Total Donations

Again, a similiar regression model to that of I and II is assumed, except that the dependent variable Y_3 is now:

Y_3 = Total dollar donations from zip area (TOTDON - 28)

Model III is the most general of the three models in that any variable that influenced total dollar yield could enter into the analysis. This model sought to identify "heavy giver" areas from an overall standpoint. As with Model II, total donations from a zip code could be identical for a variety of combinations of variable levels. Attention focuses on prediction rather than clear description of meaningful factors. Since the zip code populations differ considerably, the variables that reflect population size naturally enter the analysis first.

Independent variables that tend to correlate with the dependent variable but did not enter the analysis were as follows:

UNDER3 (2) - .891	GRTR15 (6) - .846
BTWN35 (3) - .835	TOTTAX (13) - .869
BTN510 (4) - .854	PTOV15 (20) - .514
BN1015 (5) - .872	PCTINT (24) - .546

NOTE: For an explanation of the above variables see Table IV, p. 39-42.

The summary statistics for Model III are shown in Table X.

TABLE X

SUMMARY STATISTICS - MODEL III

VARIABLE	MR	RSQ	INC RSQ	BETA COEFFICIENT	STANDARD ERROR	B/B_S F RATIO	
Constant							
INTFAM (11)	.9492	.9009	.9009	+ 4.485	.392	11.44*	
AGI (8)	.9553	.9127	.0118	- .211	.019	-11.11*	
AVGDIV (23)	.9577	.9171	.0045	+2602.381	590.168	4.41*	
DEPEXP (7)	.9633	.9279	.0108	+ 1.110	.113	9.82*	302.284
TOTFAM (1)	.9760	.9525	.0246	- .575	.150	- 3.83*	P<.0001
DIVFAM (9)	.9790	.9584	.0059	+ 5.083	1.066	4.77*	
INTDOL (12)	.9814	.9631	.0046	- 1.387	.36	- 3.85*	
AVGTAX (26)	.9823	.9649	.0018	- 596.742	279.030	- 2.14*	

* Significance
** Variables appear in the order in which they entered the analysis using a step-wise regression

Model III accounted for 96.5% of the variance in the dependent variable and included in interesting array of independent variables, as evidenced by Table X. Variables representing income sources, both primary and

evidenced by Table X. Variables representing income sources, both primary and secondary, and income amounts, represented a majority of the model. Variables indicating population figures representing total families in a zip (TOTFAM - 1), and number of dependent exemptions (DEPEXP - 7) comprised the rest of the model. A major point of emphasis is that 90% of the variance in total donations was explained by one dependent variable, dollar amount of interest received by families in a zip, (INTFAM - 11). The additional variables moved the total explained value up 6.5% to 96.5 percent.

In summary, the total donations level for an area appears to be primarily explainable by

- Large population
- High total income
- Many families having interest and dividend income

Model III can be equated as follows:

$$Y_3 = 222.87 + 4.48\ x_{11} - .21\ x_8 + 2\ 602.38\ x_{23} + 1.11\ x_7 - .57\ x_1 + 5.08\ x_9 - 1.37\ x_{12} - 596.74\ x_{26} + E$$

where

Y_3 = Total dollar donations in sample zip

x_{11} = Number of families in zip code claiming interest in sample zip

x_8 = Total adjusted gross income in sample zip

x_{23} = Average dollar amount of dividends per family in sample zip

x_7 = Total number of dependents claimed by families in sample zip

x_1 = Total number of families in sample zip code

x_9 = Total number of families receiving dividends in the sample zip code

x_{12} = Total dollar amount of interest claimed by families in a sample zip code

x_{26} = Average income tax paid per family in the sample zip code

E = Error term

A disturbing factor in the analysis was the presence of negative coefficients in both Model I (PTOV15 - 20, percent over $15,000) and in Model III, (AGI - 8; TOTFAM - 1; INTDOL - 12; and AVGTAX - 26.) Apriori reasoning would seem to indicate that these coefficients would be positive reflecting that large total dollar amounts of donations should come from areas of greater population and economic affluence.

A probable explanation for the negative coefficient is that existence of other variables not in the analysis that have a negative association on both the particular independent variable and the dependent variable. These unknown variables manifest their influence spuriously through the independent variables[1].

Multicollinearity appeared to be present in the multiple regression computation as there is a high degree of correlation among a number of the independent variables themselves. Note the correlation matrix on pages 46 and 47. This condition reduces the efficiency of the estimates for the regression slope parameters, but the efficiency of forecasts of Y, is unaffected by the correlation between the independent variables.[2] Any interpretation of marginal relationships bewteen a given variable and the total donations should be avoided. However, this weakness does not hamper the predictive power of the total equation.

In reviewing the models presented in the foregoing analysis, the original premise that total donations is a function of both the average dollar amount donated and the percent of the population that donates to the Tuberculosis Association appears to be supported. Simply stated, Model III is a combination of Models I and II. Figure 4 diagrams this relationship.

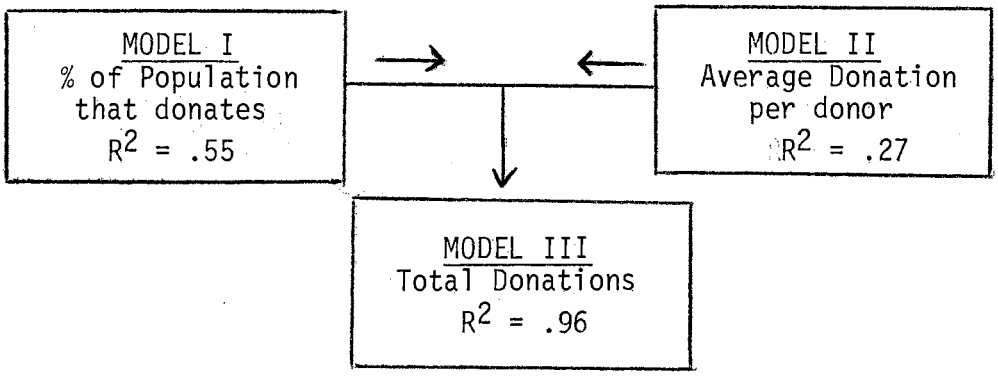


FIGURE 4. Relationship of Regression Models I, II, and III

Figure 4 shows that total donations (Model III) is a factor of both Models I and II, which can be thought of as the general or summary model. An understanding of all three models and their relationship is the key to the analysis.

FOOTNOTES

[1]Doyle L. Weiss, "Determinants of Market Share," Journal of Marketing Research, (August 1968), p. 293.

[2]William F. Massy, "Statistical Analysis of Relations between Variables," in David A Aaker, Multivariate Analysis in Marketing: Theory and Practice, p. 35.

CHAPTER VI

SUMMARY AND FINDINGS

Summary

The preceeding study attempted to apply proven marketing research techniques to an area previously almost totally ignored by marketers, that being charitable organizations. Statistical techniques were applied to a mass of data in order to attempt to segment a heretofore undifferentiated market approach to fund raising activities.

Charitable organizations as a whole, have found an increasing amount of competition for the "donation dollar" from all segments of our society, such as new charities, revised tax laws, inflation, increased personal spending, and higher costs of operating the contributory organization, just to name a few.

Because organizations, whether they are profit-motivated or charitable, are all basically similiar in structure, it would appear logical that concepts and techniques proven successful to a profit-oriented firm could be successfully applied to a foundation or charity. Little research has been done to date to either uphold or invalidate this line of reasoning. In the field of marketing however, there is a growing interest by some to reach out with marketing tools and techniques to encompass organizations previously excluded from consideration.

When the Department of Administrative Sciences at Oklahoma State University was contacted by the Oklahoma Tuberculosis Association in hopes that an interdisciplinary study could be conducted to assist them in improving their efficiency and service to the citizens of Oklahoma, one of the subject areas considered pertinent was a market segmentation study. Their method of solicitation, record system, and history as a charitable organization, all seemed to indicate that a study of this nature would be benefical to improving their operation. Many hours of literature review, data gathering and information processing were undertaken in order to gain meaningful results for the study. A number of problems were encountered, mainly in the areas of data gathering from the Tuberculosis Association's records, and the unavailability of Fifth Count census information at the time the study was undertaken. The results of the study, although being of a "first cut" nature with regards to the total scope of the problem, uncovered a wealth of information of value to the charity's operation, and has paved the way toward additional study in this area.

The Research Findings

The study focused attention on three basic models of donor behavior. Each model included a particular combination of economic variables. From the models, a variety of decision implications can be drawn. But first, the models are summarized.

Model I sought to explain the variability in "percent of donors" across zip code areas. The results suggest two basic population segments of special interest. One is the family in a higher social class and

income level. The other is the somewhat older family on a fixed income (perhaps retired). A relatively large amount of the variability in the dependent variable was explained (55.4 percent).

Model II examined the data for information regarding the size of the donation per donor. The key variable in the model was high income. If a person has made the decision to donate, a high income will be reflected in the donation. However, this model had a very low degree of explanation (27.6 percent). Its predictive power is quite suspect.

The final model was developed primarily for prediction of total donations in an area. Although the variables in the model reflect key dimensions of the problem, their interrelationships confuse the interpretation of specific relationships. The variables indicate area strength as coming from populus areas with high income from interest and dividend sources. The level of explanation was 96.5 percent which indicates a strong predictive tool.

The results of the analysis have given insight into what socioeconomic factors are indicative of the so-called "heavy giver," and where he may be found. In order for the Tuberculosis Association to remain viable and in competition with all the other organizations soliciting donations from the public, it must seek to determine how it can better and most efficiently serve its publics, both from the standpoint of maximizing returns and also minimizing the costs of obtaining these donations. It is felt that the information contained in the study will aid those involved with the Tuberculosis Association in broadening the scope and knowledge of its market for charitable contributions.

CHAPTER VII

DECISION IMPLICATIONS AND SUGGESTIONS FOR FURTHER STUDY

Decision Implications

This study, although attempting to apply marketing techniques to an organizational area almost totally without prior market research, provided a great deal of valuable and pertinent information. A great deal of insight into how a large charity functions at the state level was gained, and a very close working relationship was fostered between those involved in the study and the staff of the Oklahoma Tuberculosis Association.

The analysis of data did identify a number of key variables that can help identify and explain the characteristics of those people who donate to the O.T.B.R.D.A. In terms of methodology, it was shown that data transformations sometime allows for a better "fit" of the data, as was the case in this analysis.

With the aid of the various models developed in the study, an organization like the O.T.B.R.D.A. can develop a more definitive marketing strategy than heretofore available. For example, Model I, which attempts to define the percentage of a zip that are donors, can be used to scan a

list of the entire state's zip codes, and decide which ones a direct mail campaign should be directed toward instead of simply a blanket mailing to the entire state.

Model II will aid in the determination or identification of the large donor. This will enable an organization to tailor its fund raising campaign at various donor levels. In other words, a zip code that shows a large proportion of heavy givers might be best contacted through personalized mailing with a telephone call follow-up or other productive effort that has shown strong results with this type of giver.

Model III being of a generalized nature, probably would be most useful as a normative measuring device which would allow an organization to predict the amount of funds that should be generated from a given zip code, and would enable them to judge the overall effects of their fund raising campaign.

Perhaps the best way to accentuate the use of the results would be to pose questions whose solutions grow from the models.

Key questions that might be asked are:

1. Suppose I have the opportunity to secure mailing lists from various organization memberships, magazine subscriptions, etc. Are there general criteria by which I could assess whether the list would likely yield donations to the Oklahoma Tuberculosis Association?

2. Solicitation mailings have been sent to a wide range of groups, areas, etc. Are there some names to which I should send follow-up reminders?

3. I would like to build my donor list by the use of saturation mailing to "occupant." Are there some areas better than others to which I should send material?

4. A control device to assess how well a given area has been canvassed would be of value. Can I examine the donation files and identify areas that yield considerably less returns than should be the case?

5. Many individuals in my files have not donated in X years. Should I more readily drop individuals from some areas than from others?

6. Suppose I would like to encourage higher donations from current donors. Are there some individuals who would be more receptive to raising their donation level than others?

All of the above questions could be at least partially answered by use of the models in this study. Data for the independent variables could be inserted in the equations for the various models and predictions made on the dependent variables: percent donors, donation level per donor, and total donations.

As an example of what could be done using the Models, assume the following hypothetical situation. Suppose we want to predict the total amount of funds that should be generated from a zip code, based on our study. Using Model III and obtaining data from the zip code in question, we could arrive at a predicted level of income (donation) from the area. Additional applications using Model I and II are shown in appendix.

Example, Model III:

Using the following data from zip code 12345:

437 - families claiming interest, (X_{11})

$8,715 - total A.G.I. reported for zip, (X_8) (in 000's of dollars)
$ .11 - average amount of dividends received per family, (X_{23}) (in 000's of dollars)
1363 - total number of dependents claimed, (X_7)
1465 - total number of families in zip, (X_1)
111 - total number of families receiving dividends, (X_9)
$211 - total interest received by all families in zip, (X_{12}) (in 000's of dollars)
$ .85 - average income tax paid per family, (X_{26}) (in 000's of dollars)

and Model III

$$Y_3 = 222.88 + 4.48\ x_{11} - .21\ x_8 + 2602.38\ x_{23} + 1.11\ x_7 - .57\ x_1 + 5.08\ x_9 - 1.37\ x_{12} - 596.74\ x_{26} + E$$

the computation yields:

$$Y_3 = 222.88 + 4.48\ (437) - .21\ (8{,}715) + 2602.38\ (.11) + 1.11\ (1363) - .57\ (1465) + 5.08\ (111) - 1.37\ (211) - 596.74\ (.85)$$

$$Y_3 = \$1{,}082.28$$

$1,082.28 - Total dollar amount of donations expected from zip code 12345

The mathematical routine embodied in the Models would be best utilized on a high speed electronic computer which could scan hundreds or even thousands of zip codes to predict what results should be obtained. Armed with an analysis such as this, a mail-oriented fund raising organization like the O.T.B.R.D.A. would be better able to serve its publics. They would be able to pinpoint those zips which appear to be most productive and modify their mail campaign to achieve a saturation mailing, follow-up mailing, telephone solicitations, or other effective means of contacting people. The net result is that the group can develop a logical plan and goals once the market has been defined. As the fund drive progresses, a comparison can be made to check actual versus predicted results, and

comparison can be made to check actual versus predicted results, and adjustments to the plan, if needed, can be made. An analogy may be drawn to a fund-raising group starting a drive and a ship leaving port. How many ships leave port without the captain knowing the route and the destination, but how many organizations really know to whom they are or should be appealing and what a realistic goal should be.

It has been demonstrated that data collection and analysis, in terms of zip coding can be an efficient and practical means of market segmentation of those organizations who market by mail or compile records by zip code. The wealth of data available through the 1970 census and related data tabulations, allows the researchers to define market segments just a bit more precisely. This study attempted to combine seasoned marketing research techniques, a new store of statistical information, and an organizational setting, until recently disregarded by research. Since the data on economic analysis is dated 1969, the organization must continually update its analysis. However, given the procedure outlined in this paper, no extreme difficulties should be met.

To the extent that valuable information was gained, relative to our problem, the study was a success. To the extent that this study suggested new topics of research in the areas of non-business marketing, it was also a valuable undertaking.

Suggestions for Further Study

Census Data

The study utilized data from the Internal Revenue Service. A wealth of demographic and socioeconomic data can be obtained from the 1970 Census - Fifth Count Summary Tapes, which identify or group data by zip code. This tape was not available at the time of the study. However, a combination of data from both sources should give very significant results in a follow-up analysis.

O.T.B.R.D.A. Records

The Tuberculosis Association's records were mentioned as having been split into the "General" file and the "Special" file. Data from this study came only from the "General" file. Inclusion of information from the "Special" file would possibly closely identify the high donor. This data is not tabulated by zip code and would be difficult and time consuming to analyze, but would give a broader look at the donation trends.

E.D.P. Coding of Records

Because of the voluminous amount of records which are kept by the Association, together with their repeated use, it is suggested that a feasibility study could be undertaken to determine whether E.D.P., electronic data processing, would be more efficient in storing donor information than the present methods. The transfer of addresses could

be handled by computer based typewriters, and further studies such as the Zip Code Analysis presented in this paper would be made easier by computer based records.

SELECTED BIBLOGRAPHY

Aaker, David A. "Multivariate Analysis in Marketing: Theory and Practice, Belmont, California: Wadsworth Publishing Company, 1971, Part I.

Bass, Frank M., Douglas J. Tigert, and Ronald Lonsdale. "Market Segmentation: Group versus Individual Behavior." Journal of Marketing Research. Volume 5 (August 1968), p. 264-270.

Baier, Martin. "Zip Code - New Tool for Marketers." Harvard Business Review, (January - February 1967), p. 136-140.

Borden, N. H. and C. W. Smith. "An Appraisal of Census Programs for Marketing Uses." Journal of Marketing, (April 1954), p. 331-360.

Brown, Dr. George. "Handier and Dandier Data for Marketers." Sales Management, (December 1970), p. 20-23.

Brown, Dr. George. "Putting the 1970 Census to Work for Marketing." Business Management, (January 1971), p. 21-26.

Dixon W. J. BMD - Biomedical Computer Programs, Los Angeles, California: University of California Press, 1971, p.233-250.

Eckler, A. Ross. "Profit from 1970 Census Data." Harvard Business Review (July - August, 1970), p. 4-16.

Heyman, Timothy. "Use of Census Data in Interregional Marketing." Sloan Management Review, (Volume 12, 1971), p. 17-31.

Kotler, Phillip, and Sidney Levy. "Broadening the Concept of Marketing." Journal of Marketing, Volume 33 (January 1969), p. 10-15.

Mindak, William A. and Malcolm Bybee. "Marketing's Application to Fund Raising." Journal of Marketing, Volume 35 (July 1971), p. 13-18.

Nauman, Mack. "Multiple Regression Analysis of Outboard Motor Sales." Marketing 5223 - Spring 1972, Oklahoma State University.

Neilsen, Arthur C. Jr. "The Census - Essential Tool for Marketing." American Marketing Association Spring Conference, (April 1970).

Taeuber, Conrad. "Census Data Makes Sense." American Marketing Association, Greensboro, North Carolina, (Fall 1972).

United States Department of Commerce. "Use of Demographic Data in Marketing Planning." September, 1971.

United States Post Office. Zip Code Directory of Oklahoma, 1972.

United States Internal Revenue Service, Federal Individual Income Tax Return Data for Each 5-Digit Zip Code Area in Oklahoma: 1969 Internal Revenue Service, Washington D. C. 1971.

Weiss, Doyle L. "Determinants of Market Share." Journal of Marketing Research, Volume V (August 1968) p. 290-295.

"APPENDIXES"

APPENDIX A

O.T.B.R.D.A. - MARKETING STUDY

I. CENSUS BUREAU INFORMATION

A. Telephone Numbers

1. Public Information Office (printed matter)
   a. 301 763-7273 (Washington D. C.) - Mr. Melke, Director

2. Computer Data Service and Information
   a. 301 763-5002 - Nellie Fay Harris
   b. 301 763-5266 or 5267 - Marshall Turner

B. Count Five (5) Availability Information - Sources that have Count Five (5) summary tapes

1. University of California - Berkley

2. I.R.S. - Oak Park, Michigan

3. Kansas City Information Development Systems
   3430 Broadway
   Kansas City, Missouri 64111
   c/o Mr. Jacob Ruff

4. Cooper Communities Inc.
   Bella Vista, Arkansas 72712
   c/o Tom Seay

5. User's Service Staff
   Bureau of the Census
   Washington, D. C. 20233
   Cost: $70*

   * NOTE: Can purchase a tape from this office, printouts are not available, and cannot be obtained on a loan basis.

6. Oklahoma State University
   University Computing Center
   Stillwater, Oklahoma

   NOTE: Oklahoma State University has Count Five Summary tapes for the following:

   3-Digit - Entire United States
   5-Digit - State of Oklahoma and surrounding area (73 + area)

C. Printed Matter - Address

1. Data Access and Use Library
   Data Users Service Office
   Bureau of the Census
   Washington, D. C. 20233

2. District Field Office
   United States Department of Commerce
   1110 Commerce Street
   Dallas, Texas 75202

II. OKLAHOMA TUBERCULOSIS ASSOCIATION INFORMATION

A. Oklahoma City
   Mr. Ralph Morgan
   Secretary: Mrs. Ruth Emerson
   2442 North Walnut - P. O. Box 53303
   Oklahoma City, Oklahoma 73105
   405 524-8471

B. Tulsa - Lakes Area Tuberculosis Association
   Mr. George Martin
   Secretary: Mrs. Knight
   808 S. Peoria
   Tulsa, Oklahoma
   918 584-4238

APPENDIX B

MODEL APPLICATION

(Sample Problem)

Zip Code 74017

Model I

$$\text{Per cent Donors} = -.096 + .632X_{23} + .345X_{14} - .249X_{18}$$

$$= -.096 + .632(292) + .345(.035) - .249(.049) = .18(18\%)$$

Model II

$$\text{Average Donation per Donor} = 2.22 + 3.435X_{18}$$

$$= 2.22 + 3.435(.049) = \$2.39$$

VITA

Paul Jeffrey Scanlon

Candidate for the Degree of

Master of Business Administration

Thesis: Market Segmentation Through Zip Code Analysis - A Case Study

Major Field: Business Administration

Biographical:

Personal Data: Born in New Haven, Connecticut, May 8, 1946; the son of Mr. and Mrs. Joseph J. Scanlon

Education: Graduated from Shelton High School, Shelton, Connecticut in June 1964; received the Associate of Science degree from Oklahoma State University in May 1968, with a major in Fire Protection; received the Bachelor of Science degree from Oklahoma State University in May 1969, with a major in Occupational Education; received the Master of Science degree from Oklahoma State University in January 1971, with a major in Occupational Education; completed requirements for the Master of Business Administration degree at Oklahoma State University in May 1974.

Professional Experience: Firefighter and Ambulance Medic, Stillwater, Oklahoma, Part-time and summers, 1965-1973; Professional Sales, Gibson's Products Company, Stillwater, Oklahoma, Summer of 1968; Graduate Teaching Assistant, Oklahoma State University, Stillwater, Oklahoma, 1969-1973; Account Executive, duPont-Walston, Inc., New York, New York, 1973; Account Executive, Merrill Lynch, Pierce, Fenner and Smith, New York, New York, 1974.

Name: Paul J. Scanlon Date of Degree: May 1974

Institution: Oklahoma State University

Location: Stillwater, Oklahoma

Title of Study: Market Segmentation Through Zip Code Analysis - A Case Study

Pages in Study: 74 Candidate for Degree of Master of Business Administration

Major Field: Business Administration

Purpose of Study: Marketing Researchers have applied their tools and techniques very efficiently to traditional profit-motivated organizations. Non-business or charitable organizations, on the other hand, have been the focus of only limited amounts of research. In order to further the knowledge of marketing in this previously ignored area of organizational research, this study focused on the problem of market identification and segmentation for the Oklahoma Tuberculosis and Respiratory Disease Association. The study focused on the combination of proven statistical techniques, available research data broken down by zip code, and the historical records of the Tuberculosis Association. It was the intention of the analysis to more accurately define and predict those zip codes that would be most likely to be "heavy giver" areas. Thus, this study summarizes the field research and computer analysis of data compiled from the Oklahoma Chapter of the Tuberculosis and Respiratory Disease Association.

Findings and Conclusion: The study established the fact that statistical data complied by zip code can be a very efficient means of analyzing a market that is characterized by a mail solicitation approach. These areas can be identified as to the potential of donations based on economic characteristics of the population. The availability of data sources and statistical techniques indicate that groups such as the Tuberculosis Association should profit greatly by segmenting and identifying their donor markets.

ADVISER'S APPROVAL Stephen J Miller