INFORMATION PROCESSING

Applications
in the Social and
Behavioral Sciences

WILLIAM I. DAVISSON

INFORMATION
PROCESSING
applications in the social
and behavioral sciences

William I. Davisson
University of Notre Dame

PREFACE

This book is designed for the social and behavioral scientist. Specifically, it is for those who desire to utilize the computer in their work. The book assumes no mathematical prerequisite.

I wish to thank Dr. Harry E. Pople, Jr., of the University of Pittsburgh, for his constructive advice in developing a user oriented book. Dr. Leon Winslow of the University of Notre Dame has spent considerable time reviewing the work and helping to develop the idea of the computer application approach used. The staff of the Social Science Training and Research Laboratory of the University of Notre Dame have contributed generously of their time and aid in developing the manuscript. Individual contributers are indicated at the chapter or section contributed.

CONTENTS

vii

Introduction

This book has been written as a guide for social science and behavioral science scholars who desire to utilize the computer in analyzing and evaluating data. The book is not designed as a FORTRAN language book; although a FORTRAN review for social science readers is included as an appendix. However, it is strongly suggested that persons using this book supplement it with the manual of FORTRAN for the computer available to the individual.

FORTRAN is not an end in itself; it is only a means to an end. It is a language that allows instructions to be written to direct computer operations. The end result is the analysis and evaluation of data. FORTRAN is only useful to the scholar or researcher who has a problem which may be efficiently solved by using the computer.

The book is divided into three parts. Part I is a statement of a generalized social science methodology, or model analysis. The second chapter of the part relates the model analysis to computer analysis or computer simulation.

Part II of the book (Chapters 3–9) considers the dual problems of data and computer utilization of data files. Each data file represents a unique set that presents its own problems for the researcher. This section sets the stage for the application of computer methodology and technology to research problems in the social and behavioral sciences. Part II concludes with a discussion of standardized program procedures (canned programs), and the ways in which they may be useful to the researcher in evaluating data (Chapter 9).

Part III of the study considers computer technology together with social and behavioral science methodology in solving research problems in the social and behavioral sciences. In short, Part III considers the application of the computer to the social and behavioral sciences.

This section employs a "case study" approach. Chapters 10–15 are composed of separate problems that the researcher or scholar is likely to encounter. Each of these chapters poses a typical computer problem for a discipline; indicates the data set, the methodology to be used in developing a solution, and the nature of the solution. In instances where it will increase the understanding of the process and the solution, the actual programs are shown and explained.

one
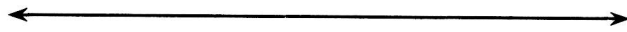
INFORMATION PROCESSING
AND COMPUTER USE

1

Social Science Computer Use

What is analysis?

Analysis of any problem or issue precedes any possible computer or other information processing or statistical analysis. In a broad sense, social science analysis consists of two steps and involves the use of two kinds of tools: *logic* and *statistics* (statistical inference).

Evaluation of social science problems usually focuses on the policy implications thereof. The state legislature will require the evaluation of the effect of a particular program in order to decide whether to increase the spending on the program or to eliminate the program. The federal government may desire an evaluation of the expenditures on the Poverty Program in the United States, or on the Peace Corps, in order to decide whether to increase or decrease the expenditures thereon. Somebody has to make a decision whether to give federal and state monies to a given city in the United States to help with the cost of an urban redevelopment or sewage control project. The unique aspects of social science analysis are that the results of the analysis almost always point (in one way or another) at immediate policy decisions by someone, and the social scientist never works with a test-tube situation. The social scientist deals with a situation in which he himself is a *part* of the decisions or questions that he is testing. Rarely is the physical scientist part of the experiment that he is undertaking — e.g., he is not part of the atom *being exploded*. In a very real sense, for anyone familiar with government, business, or social science decisions, the person doing the evaluating may be the same person or thing being evaluated.

One of the more difficult elements of social science analysis is to formulate a specific question (policy) that may be analyzed. Rarely does one ever find a nice, precise, social or economic question standing around all by itself in a form that is immediately susceptible to testing.

A primary problem of social science analysis, therefore, is to break up the overall issue into questions that may be answered. In short, the big problem in the social sciences is to formulate these questions. The second problem of the analysis is to be able to give statistical or quantifiable answers to the questions.

The first step of the problem is usually termed "developing or formulating the hypothesis — i.e., the question — to be answered." The second step involves testing the hypothesis. Thus, there are two broad sets of tools in the social scientist's kit: logic and statistics. Let us summarize the question of analysis and determine where

logic and statistics fit into this overall picture. We may use the schematic outline below to illustrate the point:

I. *Logic:*
   A. Statement of the overall policy or issue.
   B. Delineation of assumptions: What component elements of the problem will not be discussed?
      These elements are considered to be the things that will be taken for granted, or at least will be omitted from the discussion. Assumptions are so-called negative elements — things that will not be considered.
   C. Focus of the overall issue into its component parts with the eventual definition into a specific question. The answer to this question will measure the overall policy or objective. *This is the hypothesis.*

II. *Statistics* (Methodology):
   A. Methodology.
      1. The Information Source: What statistical or arithmetic information is available to be used in answering the question?
      2. The Selection of the Sample: Shall all the available information be used in answering the question, or shall less than all the available information be used?
      3. The Calculations: The process of converting the raw statistical (arithmetic) information into a specific numeric answer to the hypothesis.
   B. Verification of Information, Sample, Methodology: Is the analysis proceeding in a manner to actually do what it is supposed to do?
   C. Conclusions: Let us take this schematic outline into the real world of social science analysis and see exactly how it is applied.

Urban renewal

Urban renewal[1] was one of the more widely discussed and argued domestic issues of the 1950's and early 1960's, and still is today. Urban renewal primarily involves removal of defined substandard (slum) housing and replacement with newly constructed housing sponsored by federal aid. Aside from the groups directly affected, many different groups get into the act: realtors, contractors, political scientists, sociologists, city planners, architects, local housing authorities, and several federal agencies. The critics of the so-called urban renewal are almost as numerous as the number of persons involved.

Civil rights groups have protested that urban renewal is "Negro removal." Social workers suggest that relocating families from areas to be renewed means simply moving the slums from one part of the city to another. Economists have noted that urban renewal programs substitute high- or middle-income (priced) rental units for low-income (priced) rental units, and that the basic problem is poverty, not housing renewal. The federal agencies and the cities hope that urban renewal will halt the continual decline in property values in the central city.

At this point, no specific problem of urban renewal has been brought into focus. Before a problem can be analyzed, that problem — or the relevant segment of the

[1] See Chapter 10.

problem — must be sharply defined. When the problem is defined, then it may be analyzed to determine the nature of the problem, and how an answer may be developed. The first point in analyzing the "problem" of urban renewal, therefore, is to define precisely what the problem is. This in itself is often a real problem. The Urban Renewal Program, as with many other programs, was a matter first of federal policy supplemented by local government policy. In order to appropriately criticize urban renewal, one must turn to the instruments of public policy (statutes) that established the program to determine what the major objective or objectives of the program were.

The main goal of public policy here appears to have been the decrease in the stock of substandard and overcrowded housing in a city. If we accept for the moment that the major purpose of the Urban Renewal Program was to decrease the stock of substandard and overcrowded housing, we have defined our problem. We then intend to analyze the efficacy of a major federal expenditure policy program. Were the expenditures justified as measured by the impact of the expenditures on the housing supply? In short, what were the actual results of major urban renewal efforts on the changes in the population and housing characteristics of cities?

Without reference to a computer, to data processing, or to programming, how would you, as a social science student, *analyze* that problem? What would you have to do, and how would you go about developing an answer to the question posed? We do not contend that this is the only–or even the most important question surrounding urban renewal. We only ask, if you as a social scientist were faced with answering the problem posed, how would you do it? If you had only a pencil and piece of paper, could you answer that problem? If so, how?

In considering our question, let us start at the beginning and determine how we arrived at it. It is important to understand that from the viewpoint of social science analysis, once one has been able to pose the question, one has already progressed in the analysis.

The basic policy involved is the federal policy of providing aid funds to local governments for various kinds of "urban redevelopment." In other words, the federal government gives money to cities for a given purpose. The logical question is then: Has the policy (urban redevelopment) been successful? The question posed above indicated that the success (or failure) of the policy was to be measured in terms of population and housing characteristics of cities. Our problem here is to determine just how the problem was defined to this point. In order to do this step by step, let us again take the broadest possible question: Has urban redevelopment (expenditures by the federal government) been successful?

Immediately one is faced with the problem of definition: What is success? In order to answer this question, it is necessary to determine what was the major intent of the policy. The urban redevelopment expenditures are made according to federal law. The first step, therefore, is to examine the law itself to see what it says concerning the avowed purpose of federal expenditures. If the law itself says something concerning the purpose of the expenditures, this may be a help. However, in a democratic society, a law is nearly always a compromise — a compromise among those who wanted the law in a stronger form, those who wanted the law in a weaker form, those who wanted no law on the subject at all, those who wanted a different law but took this one, and so forth.

Thus, in order to determine of what the existing law was a compromise, it is necessary to determine who the interested parties to the issue were. In order to do this, it is necessary to go to the Committees (House and Senate) bills, speeches, hearings, and presentations of other interested groups (lobbyists). In this way, the

social scientist may gather some idea concerning the various groups who were interested in and working toward or against some kind of urban redevelopment bill. In this way, one may gain an impression of the manner in which the existing federal enabling statute (as signed by the President into law) was and is a compromise. This kind of preliminary analysis is essential in order that the researcher may be able to have some understanding of the nature of the policy issue involved, and therefore, some understanding for the type of yardstick by which to measure the success or failure of the law.

Success of the federal policy of urban redevelopment, therefore, in some elemental sense is really a measure of what the various contending groups would measure as success:

A. Success to the city might involve improvement of housing conditions, taxable property values, and overall city beautification. Sucess to the elected city officials might involve being reelected.
B. Success to realtors and property owners might be improvements in property values.
C. Success to businessmen might be an improvement in general business and sales in the center of the city occasioned by the results of redevelopment expenditures.
D. Success to civil rights groups might include improved (and/or subsidized) housing for the ethnic groups.
E. Success for the so-called relatively uninterested but esthetically inclined city planners might be a more ordered and beautiful city.
F. Success for the so-called preservation societies might be defeat of the program because it would tear down some old existing landmark that is in some way identified with the history of the city.
G. Success for the contractor and the federal politician might be the full employment attendant on the federal expenditures.

Thus, the question of the success of a federal policy such as expenditures on urban redevelopment involves a host of questions, and in one way or another means that the researcher must concern himself with the parties contending the issue.

A clue, however, to the nature of the problem is found in the fact that the federal policy of urban renewal was implemented by means of federal monetary expenditures to cities for a given purpose. Another clue to the nature of the issue is that in most of the presentations of the parties involved, the question of *slum clearance* was involved — i.e., overcrowded and substandard housing.

Since the major aim of the urban renewal expenditures has been in the direction of slum clearance, including primary emphasis on removal of substandard and overcrowded housing, it may be quite logical to determine the success or failure of the policy in terms of the degree to which the federal expenditures alleviated conditions of overcrowded and substandard housing. Thus, this quite theoretical analysis has brought us through a maze of political and other obstacles to a clear statement of the aim of this study in measuring the success or failure of urban renewal: *What were the actual results of major urban renewal efforts on the changes in the population and housing characteristics of cities?*

At this point, the analysis (breakdown of the problem or issue into manageable component parts) has proceeded to the formulation of the question and the development of a hypothesis. The testing of the hypothesis with the resultant answer will, accurately or not, then answer the question asked. Three things still need to be done:

first, the social scientist must develop a methodology that will allow him to answer the question posed; *second*, he must determine whether the approach will in fact provide the desired answer; *third*, the social scientist must finally evaluate how adequately the question posed and answer developed bears upon the original and basic question. In other words, just how precisely does an answer to the question of housing and population bear upon the question of the success of the federal policy of urban renewal? Thus, in broad scale, the social scientist's analysis of the problem of urban renewal has been divided into two parts: (1) the definition of the basic issue of expenditures on urban renewal to a manageable policy issue; (2) the development of a methodology to answer the issue defined.

In order to answer the question posed, it is necessary for the social scientist to have data that will permit an answer. The question posed concerned housing and population. Immediately, then, there are three consistent and reliable sources of information that bear directly on housing and population. The decade population census figures show for each city, and for each census district in each city, detailed figures on population, ethnic groups, housing, and related matters. The second basic source of information that the researcher might utilize would be building permits — although in some cities these would not be available because of the nature of the record keeping (or non-keeping) or for other reasons. The final source of information that would bear on the question of housing, and to a certain extent on population or housing residents, would be the city or county property tax rolls. Of all these sources only the census figures may be relied on as being relatively consistent over a period of time and among various cities. Thus, without detailed personal research in the cities or urban areas involved, only the census data may be reliably used.

The determination of the cities to be used in the examination will undoubtedly turn on two factors: the size of the cities and the comparability of the cities as to geographic area, industrial-economic characteristic, general income level, and ethnic patterns in the city. In evaluating the significance or success of urban renewal, it will be most appropriate to choose cities of about the same size measured by population, with a similarity in the other factors noted.

Having selected the cities, then, two questions must be answered using the data available. The condition of the cities, measured by population characteristics and substandard and crowded housing, before the urban renewal program was initiated, must be determined. The condition of the cities, measured by population characteristics and substandard and crowded housing, must be determined after some time period has elapsed. Normally, the city selection should involve one or more cities that did not utilize urban renewal funds, and one or more cities that did utilize these funds. By statistically evaluating the changes in the population and housing condition, these data utilized will permit some kind of quantitative evaluation of the impact of urban renewal programs on population and housing conditions, *both for cities utilizing the program and for those cities not utilizing the program.* Evaluation of the data will permit some kind of conclusion on the success or failure of urban renewal insofar as population and crowded and substandard housing are concerned. It should also be clear that the definitions used of population characteristics and of *crowded* and *substandard* housing are those provided by the census data itself (another caveat.)

For the social scientist, the conclusion that one derives depends upon the accurate definition of the problem, the nature of the data used, and the limiting definitions within which the research must be undertaken. Careless analysis, or inappropriate methodology (approach to solving the problem) will not provide accurate answers. In summary, the steps (analysis) involved:

A. Statement of the overall policy and objective: the success of urban renewal.
B. Analysis (breakdown) of this into its component parts and the eventual derivation of a specific question, the answer to which will measure the success or failure of the overall policy.
C. The methodology:
   1) The information sources.
   2) The selection of the sample (cities).
   3) Statistical information desired:
      a) Before urban renewal.
      b) After urban renewal.
      c) Impact or urban renewal (or of not using the program).
D. Verification: Will the methodology answer the specific question?
E. Conclusions.

Where the computer fits

The point made above is that the primary function of the computer is to manipulate information with the intent of drawing conclusions from it. The computer may be used to handle arithmetic or alphabetic information. The computer may be used as a giant calculator to perform arithmetic manipulation of the information. It may be used as a combination clerk and file cabinet for data storage and retrieval. The computer may be used as a statistical tool in developing a predetermined and defined sample from a mass of information.

Actually, the method used to solve or evaluate the problem depends on the equipment used. If one used a desk calculator, a good deal more time is spent in insuring that you have a valid sample to work with. Without automatic data-processing equipment, it is simply not possible to handle all the information available on a subject, nor even a major portion of the information. It is necessary to extract from the data a relatively small sample — random, stratified, or otherwise. The *sample* is then evaluated by hand, and the characteristics of the whole are inferred from the sample.

Once it is decided to bring a computer into the picture, one simply uses the *entire* population — i.e., all the information — or a very large sample of the data. It is then possible to evaluate the information using statistical techniques and the computer. Once the decision is made to use the computer, less time is devoted to insuring a representative small sample, and more time is devoted to evaluating a very large sample or the entire population.

Let us review our outline indicating the steps taken in analyzing the problem to see where the computer fits.

A. Statement of the overall policy and objectives.
B. Analysis of the overall policy or issue into its component parts and the eventual breakdown and definition into a specific question. The answer given to this question or *hypothesis* shall measure the success or failure of the overall policy or issue.
C. Methodology (Statistics)
   1) The Information Source.
   2) The Selection of the Sample
      COMPUTER   INFORMATION PROCESSING
   3) The Information Calculation.