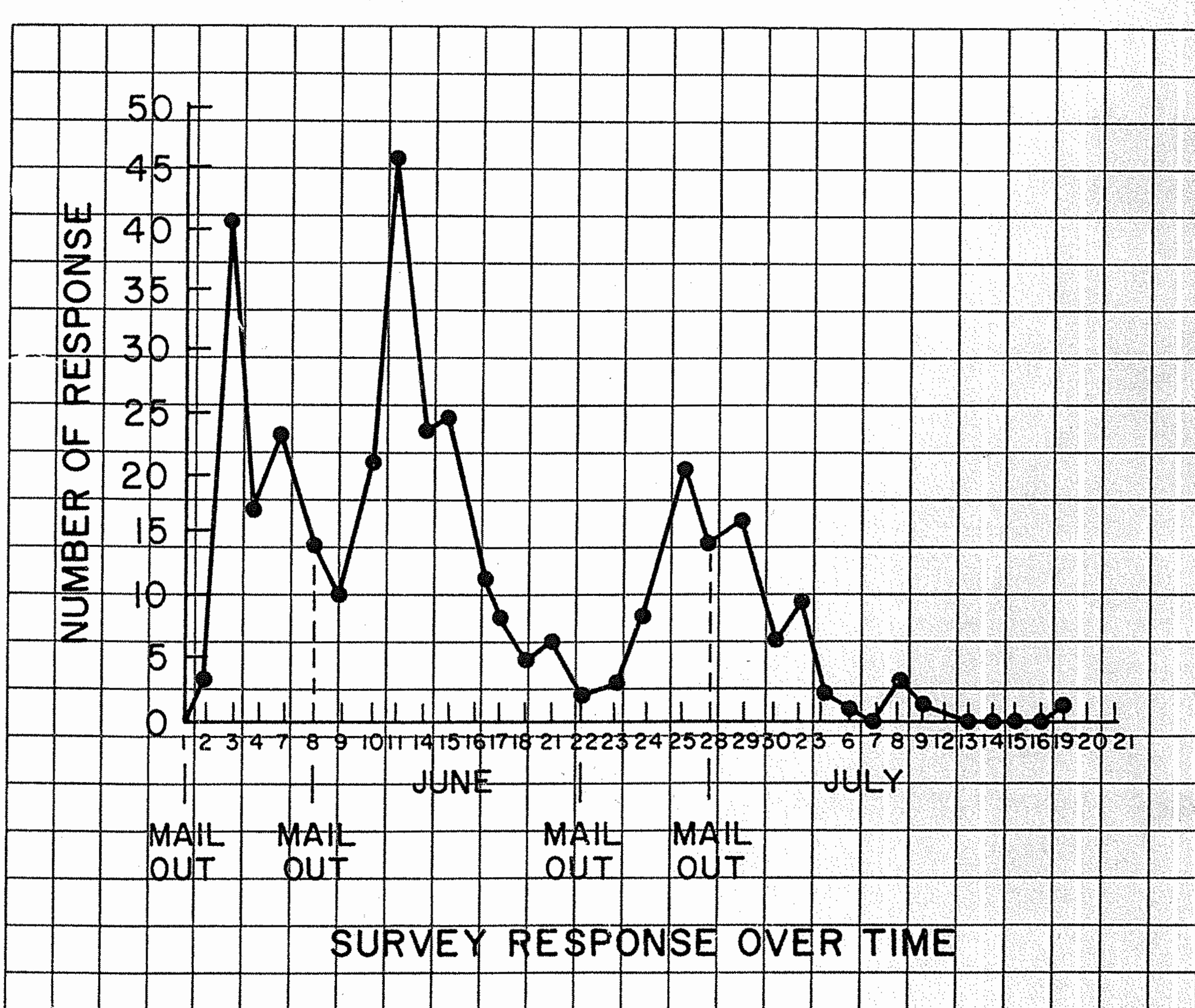
Survey Research Manual



THE UNIVERSITY OF MANITOBA
FACULTY OF ARTS

Institute for Social & Economic Research

ISBN0-9691641-5-7

SURVEY RESEARCH METHODS

(Second Edition)

by

Greg Mason
Brian Mcpherson
Derek Hum
Lance Roberts

Institute for Social and Economic Research
Faculty of Arts
The University of Manitoba
December, 1983

ACKNOWLEDGEMENTS

The many comments offered by students who attended the Survey Research Workshops offered by the Institute for Social and Economic Research are gratefully acknowledged. In addition, Candice Hammock read an earlier draft and made many useful stylistic suggestions. Don Sabourin, Statistical Methodologist at the Institute, also read the entire manuscript and pointed out the "gremlins" which inevitably arise in technical manuscripts. Elke Morrison, Cathy Watt and Grace Schuster are responsible for the typing and computer typesetting and their excellent work is much appreciated.

Finally, the financial assistance of Health and Welfare Canada, which provides much of the operating funds for the Institute is acknowledged.

PREFACE

The use of survey research methods has grown dramatically since the 1930s when the first political opinion polls were used to forecast election results. At first the record of this new measuring device was very poor, but during the postwar era, new statistical procedures, especially in the area of sampling, and new techniques of questionnaire design have considerably refined the precision of survey research. Now survey research is very common and few people in North America have never participated as respondents in a survey.

This rapid increase in the use of surveys for issues ranging from the Neilsen ratings to analyze TV viewership to social policy design raises important questions. The media routinely reports the results of most polls with little understanding as to what makes a reliable survey. Many research managers are sceptical about this methodology, convinced that the results are "softer" than data which has been filtered through an official agency. Yet many of the so-called "hard" economic data, such as unemployment and price indexes are derived from elaborate surveys. Many myths about survey research cloud effective use of this instrument. For example, many are convinced that mailout surveys are doomed to low response rates, in the order of 15 percent. Also, telephone surveys are viewed by many with great disdain and the face-to-face interview, although costly, remains the tool of choice in the area.

This manual reviews the essential concepts in scientific survey research. It begins with a discussion of the scientific method, and how survey research fits into this paradigm. The second chapter addresses the very important issue of

With any joint venture, especially one that has resulted in a manuscript that has evolved over a year, ascribing credit for the text can be difficult. Lance Roberts was primarily responsible for Chapter 1, while Derek Hum wrote Chapter 2, Brian Macpherson wrote Chapters 3 and 6, while Greg Mason is responsible for Chapters 4 and 5, the Appendices, and is editorially responsible for the entire manuscript. Although the formal description of authorship is important, it must be stressed that this work is the result of much joint collaboration and consultation among the authors.

Greg Mason, Director,
Institute for Social and Economic Research
December, 1983

CONTENTS

Chapter 1

SURVEY RESEARCH IN BROAD PERSPECTIVE

1.1 INTRODUCTION

The objective of this first chapter is to introduce a number of issues relevant to placing survey research in proper perspective. Specifically, this first section is designed to provide the reader with:

1. a clear idea of survey research as a general methodological strategy;
2. understanding whether or not survey research is appropriate for any particular problem;
3. appreciation of the strengths and weaknesses of utilizing different kinds of survey research.

As with all scientific methods, survey research works best only in certain well defined situations. Much of the misuse of survey research can be traced directly to poor conceptualization by investigators in the early stages of problem formulation.

1.2 SOCIAL SCIENTIFIC RESEARCH

In order to obtain a proper perspective of what survey research is and whether it can be useful, it is necessary to place survey research within the context of social scientific investigation. This section explains the purposes of collecting information and distinguishing how the collection of "scientific" information differs from other ways of "knowing".

ate. In this category there are those who study foreign languages that have been dead for centuries, or those who develop the frontiers of mathematics. In contrast, the practice of "applied" research is concerned with acquiring information to solve practical problems, such as turning garbage into usable heat, or conducting voter opinion polls.

Although there is something to the distinction between pure and applied research, the differences should not be exaggerated. The history of the Western World offers many examples of what were orginally mere theoretical discoveries, later turned into practical advantage. Nuclear physics, boolean algebra and even geometry are all examples of what was once useless knowledge, but which have had truly profound application in solving practical problems.

1.2.3 Methods of "Knowing" As Ways of Reducing Uncertainty

Another point that deserves mentioning is that any method of "knowing," from science to mysticism, pursued for whatever reason, is fundamentally a means of reducing uncertainty. When people find themselves in situations where they do not understand what is happening or cannot anticipate what is going to occur (both common experiences), they characteristically feel uncomfortable. For example, an important examination or application for a loan usually produces tension and anxiety. Reducing this tension requires more information by one or other method of "knowing."

1.2.4 Four Common Ways of "Knowing"

There are many particular ways by which humans gather information, and thus knowledge and certainty about themselves and their surroundings. Basically, these methods can be categorized into four groups - tradition, authority, subjectivity and science.

tions (which provide the excitement of being at the track), people turn to one method of "knowing" or another. Some follow the traditional betting methods taught them by their parents. Others hang around listening for "hot-tips" provided them by authoritative sources. Some just listen to their internal "vibes" to tell them which horse will win.

People using each of these techniques sometimes win, and sometimes lose. Important for present purposes, however, is what occurs when others attempt to learn and apply one of these techniques. What often results is a state of frustration since these methods are, on the whole, <u>non-public</u>. This is their principal flaw. There is no effective method of questioning traditional wisdoms, of challenging authorities, or "knowing" that another's intuition is really correct. In short, to employ a non-scientific method of "knowing" is to find oneself, more or less, at the mercy of the sources, even though their answers may be true. Since, any method of "knowing" runs some risk of error, the condition of being at the sole mercy of the source is an uncomfortable and an inefficient method of collecting truth in the public domain.

1.2.5 <u>Scientific</u> "<u>Knowing</u>"

The great advantage of any scientific way of "knowing" is that it resists the flaw characteristic of non-scientific methods by attempting to make the technique for arriving at conclusions publicly available. Scientific investigation attempts to perform this in two ways. First, it tries to make the connections between its theoretical ideas logically valid. Second, it attempts to demonstrate the validity of its assertions by empirical evidence; in other words, by showing our senses that they are correct.

A second source of frustration with the scientific approach occurs in the nature of its answers. Science usually cannot supply quick solutions to problems or answers to questions; rather, answers come only after a course of careful and time-consuming inquiry. Moreover, when scientific answers are provided, they rarely take the clear-cut form that many inquirers would like; in other words, scientific results are usually probabilistic in nature rather than certain.

These reasons for dissatisfaction with the scientific method are really more a comment on those who use the method rather than the approach itself. Those who wish to approach the world in "black and white," "all or nothing" terms will be disappointed, as will be all those who want instantaneous results. These understandable human desires do not, however, change the fact that the world is more "grey" than many desire. In fact, finding solutions to most human concerns takes times and effort and, rather than being a source of frustration, the proved approach of science is extremely valuable.

1.2.7 Natural Versus Social Science

The question about whether the type of scientific method practiced in the "natural" disciplines like biology and physics is strictly applicable to the "human" disciplines is a source of heated academic debate. However, there exists a substantial set of arguments that suggest that the general form of science so successfully practiced in the natural sciences should be approximated as closely as possible in the social sciences which study human behaviour. Now, of course, science is not the only way to gain insight into human habits (as poets and philosophers demonstrate), but it is one legitimate approach that carries the general advantages cited earlier. And, although the social sciences, partly

1.3 THE NATURE, STRENGTHS AND WEAKNESSES OF SURVEY RESEARCH

Survey research is perhaps the most frequently used data-collection strategy in the social sciences. Everyone, at one time or another, has been a respondent to a survey of some type and many have conducted surveys of their own. Basically, there are two forms in which surveys are conducted - questionnaires and interviews. But before the finer distinctions are explained, one must first appreciate the general characteristics of all surveys.

1.3.1 Characteristics of Survey Research

The most distinguishing feature of survey research is the fact that data are collected from representative samples of the population. Aside from a census, almost all other data collection strategies employ restricted numbers of respondents, whereas surveys may have over a thousand persons in the sample. Later in this text explicit sample selection procedures and sample size definition are covered, but for now it is sufficient to appreciate that the goal of surveys is to be representative. A second characteristic of surveys is that they seek information directly from the respondent, either by having respondents record their own responses or by having an interviewer do it for them. This direct interaction between the persons being studied and the instrument used to collect the data gives survey research results a direct, immediate tone uncommon among other methodologies. Thirdly, surveys are often conducted in "natural" settings like the home or office of the respondents. This feature distinguishes surveys from most experimental methods which are usually conducted in a laboratory and is a reason why many find the survey results less "artificial" than experimental findings.

competence, for instance. A danger does exist, however, in that unskilled interviewers can act to bias responses by placing poor interpretations on a question or prompting for a certain type of response.

1.3.3 Choosing Between Questionnaires and Interviews

There are three basic considerations involved in deciding whether to use a questionnaire or interview. The first, and usually predominant factor is cost. Since interviews involve training, wages, and travel, and questionnaires involve mostly postage and materials expenses, the expense of interviews is often prohibitively high. For example, a six question opinion poll may cost $5 an interview. If the sample size was 2,000, the cost would be $10,000. If materials and postage for each person in the sample amounts to $1 per person, the cost of the survey would be only $2,000. Of course, this does not include the cost of data processing, analysis or preparation of the report.

On the other hand, a face-to-face interview can usually be quite expensive. Interviews lasting as long as two hours are possible, whereas if a mailout questionnaire takes more than 15-20 minutes for the respondent to complete, the response rate will plummet unless special precautions are taken in the design of the survey instrument.

The reduced cost of questionnaires over surveys is obtained at a price, however, which leads to the second consideration in choosing between interviews and questionnaires - that of response rate. Response rate refers to the proportion of the sample which can be persuaded to respond to the survey. Most statistical techniques assume that the response rate is 100 percent, but, of course such rates are never achieved; researchers attempt to approximate this goal and use response rate as one rule-of-thumb indicator of the representativeness of their about their age, religion, market purchases, and the like. A second type of content often included involves "beliefs" or a person's perceptions about the facts of the matter. Third, surveys often ask questions about a person's "feelings" or emotions on some topic, or they ask about one's "opinions" or standards of action, which include ideas about what should be or could be done in a given situation. Fourth, surveys can be used to gain information about an individual's past behaviour and, very commonly, attempt to measure the reasons for people's thoughts, feelings, and actions. In such circumstances surveys ask the question "why" in an attempt to unlock the sources of behaviour.

1.3.5 Words and Deeds

Before some of the more general strengths and weaknesses of the survey research approach are reviewed, special mention is necessary, of the relationship between words and deeds. All surveys, whether questionnaires or interviews, rely on what is called "self-report" data - that is, on what people say. Admittedly, what people say is empirical data and can be used for analysis. The relevant question, however, is whether what people say, their words, is closely related to what people actually do, their deeds. This is a problematic and recurring issue that must be assessed in each survey research project. Although there are few general guidelines, one conclusion is certain: there is rarely a perfect correlation between words and deeds.

There are many reasons why such an imperfect relationship exists, but two are pronounced. First, people often just do not know, even though they are very chary to admit this fact. People may not know what they feel, or believe, or even what their opinions are. Moreover, it is the exception rather than the rule when people can truthfully report why they think, feel, or act as they do.

precisely what indicators were used to measure a variable. Finally, as previously mentioned, surveys are the best method for acquiring information about a respondent's past history (like childhood experiences) and mental motivational conditions (like attitudes, opinions, and beliefs).

With all of these advantages it is not surprising that surveys are commonly employed. As with all data collection techniques, surveys are not a flawless procedure and often the strengths in one situation become weaknesses in another. This latter situation occurs for survey research in at least two respects, the first of which concerns standardization and the second of which, concerns flexibility. As we have already mentioned, survey research is "flexible" in that is can use a variety of questions to obtain measurements about a given concept, and this is more than many methodologies can provide. However, this flexibility is limited by the format of questionnaire and interview schedules which cannot be changed while the research is in progress. This limitation can be serious if some new relevant variables or ideas are discovered in the middle of a research project and the data collection instruments cannot be changed to "effect" this new information. Like flexibility, the advantage of standardization also poses limitations on the survey technique. In order to be useful to a diverse audience, the standard form of questions posed must use "lowest common denominator" principle, and thus often appears too superficial and simplistic to validly measure the complex issues under consideration. In short, the use of standardization risks forcing "square pegs into round holes."

There are several other limitations to survey research techniques that deserve mention. First, there is the constraint that questionnaires or interviews cannot be used on certain types of populations. For instance, children and illiterate persons cannot complete questionnaires, while interviews are only use-

Issues usually involve only one concept or term and, when asked about them, many respond with declarative sentences like, "I am interested in racial conflict (or unemployment)." By contrast, research problems usually involve ideas, which, in turn, require a statement about the relationship between two (or more) concepts. Thus, people often speculate about a possible relationship between two concepts. For instance, "Does the incidence of poverty influence racial conflict?" or "How are inflation and unemployment related?" Such speculation about the relationship between variables serves the useful function of stimulating clear thought and provides a more manageable, specific focus to any investigation. The best kind of research thinking does not stop at expressing an interest in whether or not two variables are related, although research can proceed from this point. Ideally, thinking about the relationship between two concepts can become even more refined to the point of being able to specify the nature of the expected relationship between the variables. When this is accomplished an hypothesis is formed. An hypothesis is an "educated guess" about an anticipated relationship, such as "As poverty increases racial conflict increases." or "As inflation decreases, unemployment increases." Note that these are much more specific assertions than those found in issues, and because they are more specific this can be more clearly tested and definitively answered. More precisely, an hypothesis is a logical consequence of a theory and validation of the hypothesis is part of the validation of the theory.

The importance of the stage at which causal relationships are hypothesized cannot be understated. It is through these educated guesses that the survey researcher defines the purpose of the exercise, and, most importantly, proceeds to the next stage of measuring variables such as "racial conflict," and "inflation." Most often, after reflection, many common everyday terms are seen as complex and capable of a wide variety of interpretation.

1.4.3 How Misunderstanding Occurs

Appreciating this distinction between abstract and concrete levels allows us to understand what a great advantage it is to be able to communicate with concepts. With concepts it is no longer necessary for physical objects to be present in order for us to transmit a message about them. All we do is communicate the concept (usually through a spoken or written word) and it brings forth the same image in the mind of the listener or reader (assuming, of course, that all participants speak the same language).

Far more frequently than desirable, we have all had the experience of trying to communicate with another person but learning that you did not (or receive) a clear message or image. Now, how do such instances of misunderstanding occur? The most common reason is that the two participants do not share a similar understanding of terms. For example, if I asked on a survey, "How are your 'synapses'," then effective communication will not occur since, at least in this instance, we really are not speaking the same language. Of course, some do "know" the meaning of all concepts, but for most who do not we define our terms.

1.4.4 The Nature and Problems of Conceptual Definitions

Basically, there are two ways of defining any concept. The first, and most common method, is to provide what is called a conceptual definition. These are the types of definitions found in dictionaries and what they do is define a concept in terms of other concepts. The idea is that one will be able to understand the meaning of a term if it is explained using other concepts whose meaning is already known. Usually conceptual definitions state only the essential features of some phenomenon and, as such, they are abstract in nature.

To illustrate the distinction between conceptual and operational definitions, consider "intelligence." A common conceptual definition of "intelligence" is, "Intelligence is the ability to think abstractly." This defines the concept "intelligence" by using other concepts and thus qualifies as a conceptual definition of the term. Although such a definition may satisfy some sorts of curiousity, not everyone may be clear on what precisely is meant, perhaps because they do not appreciate the meaning of the terms "thinking," "ability," and "abstract." One might better inform these people by providing an operational definition of intelligence which would illustrate the term. For instance, one could say, "Intelligence is the score on The Stanford-Binet I.Q. Test." This test is something concrete that can be experienced and by using such a test one would show what intelligence means.

Clearly such demonstrations of meaning can "drop away" too much. Considerable controversy exists over whether standard I.Q. tests measure a full range of what is accepted as intelligence. For this reason several interpretations may be needed. This bias leads survey researchers to employ several measures of the same concept. By using a number of indicators or variables, errors produced by imperfect definitions can be analyzed, and hopefully, minimized. Once again this is a major advantage of a questionnaire or interview.

The reason operational definitions are so prevalent and important is that they transform fantasy into fact. Therefore, operational definitions as measures of concepts are non-existent in any tangible sense.

To summarize remember that science is interested in facts which are concrete or "real." But bare facts are uninteresting; they are collected to test ideas which are "unreal" in the sense that they exist only in our minds. First, one must have a clear picture of what the idea is and, in this regard, conceptual these procedures to survey research, we must appreciate that survey research is one set of procedures for making measurements, and in this procedure measurements are made by asking questions. So what exactly are the questions on a questionnaire or during an interview schedule? In fact, each is an operational definition of a particular concept.

Appreciating this conclusion, one very common fault in survey research designs becomes clear, - they are done because their results or answers will be "interesting." Interesting or not, questionnaires or interview schedules that take this outlook are often poor because it is often very dificult to know what concept the question is measuring; consequently, the results are meaningless. Furthermore, even if the results are interpretable (for example, to the question: What is your age?), where it is not apparent why the question was asked (that is, what other concepts are expected it to be related to), the results are not useful. This, is one of the reasons why the census, which asks large numbers of questions from great numbers of people, is underutilized (compared to its cost); most researchers do not know why the questions were asked and why some data are collected while others are excluded.

In some cases, a survey may be undertaken merely to describe a population. This can be a useful objective of surveys, but given the effort in placing it in the field, it is wasteful not to pursue a more definite program of hypothesis testing. As shall be argued below, surveys which involve the respondent in more than just a fact reporting role are more likely to have high response, and more likely to produce accurate and interesting results.

The lesson to be drawn from this is that survey research is not executed in a vacuum and cannot, by itself, usefully solve problems. To be functional, survey research is only undertaken after a lot of careful thinking has been conducted must have changed or varied in the first place; a baby must have begun crying or the unemployment rate must have dropped, etc. In other words, if things remain constant, there is nothing to explain. Now, in order to explain variation in some variable, the potential cause must itself be a variable since it is impossible for something which does not change to cause something which does. These ideas are summarized in the oft-repeated assertion that, "If there isn't any difference, it makes no difference."

When thinking with variables it is important to appreciate that "objects" or "things" are NOT variables; they are simply concepts. For instance, a chair is clearly not a variable, what is "variable" about a chair or most other things are their qualities or characteristics or dimensions. Thus chairs differ in their color, texture, shape, etc. It is precisely these qualities of objects that are measured with operational definitions. In this regard we can see why both conceptual and operational definitions are essential to the process of measurement. Conceptual definitions, among other things, specify the dimensions or qualities of the concept under consideration, while it is the task of operational definitions to formulate definitions and clear measurements of each of an object's dimensions.

1.5.2 Levels of Measurement

So far we have discussed what operational definitions are and have noted that these are the measuring instruments used to examine the values of various objects on particular variables. We now provide a somewhat fuller appreciation of the notion of measurement, for this understanding is essential to the selection of appropriate statistical techniques to analyze your data after they have been collected.

sential that the <u>rules</u> by which objects and numerals are conceived be made explicit. This is, of course, why we call the common measuring device (a yardstick) a "ruler" for it can be used to systematically attach numerals to objects. It is for this reason that we make our "rulers" out of materials, like wood or plastic, so that the measurements we make will be systematic and invariant.

Appreciating these characteristics of measurement, we should note that our effectiveness in both scientific and everyday life depends largely on our ability to <u>distinguish</u> among objects and make differential responses to them. It is of consequence, for instance, to be able to tell a poisonous from a safe mushroom, or a neurotic from a psychotic. The process of measurement is made precisely in order to distinguish between objects. Now, when we measure an object, we would ideally like to be able to portray the complexity of the object in the numeral we assign to it. Since the variables that we measure in objects differ in their complexity, it is important that we have different means of measuring different variables. This is accomplished through the application of what are called "<u>levels of measurement</u>."

There are four commonly used levels of measurement and each of these ways of measuring contains different amounts and complexity of information. These different kinds of information allow certain types of arithmetic operations to be legitimately performed. This is mentioned now because a later section of this text demonstrates that the statistics chosen to analyze data are dependent on the levels of measurement of the variables in the hypothesis. It is essential the form of survey questions (the operational measures of variables) be developed with care since this will influence the measurement level, the nature of the statistical testing, and ultimately the information provided by the survey.

ables implies, but these need not concern us here. What we can appreciate though, is that the values of ordinal variables can be compared using terms like "greater than" "less than" or "equal to." The following example of a question whose values are ordinal illustrates this is the case.

2. Pierre Trudeau is a first-rate leader.
(Please check one)

1. Strongly Agree ( )
2. Agree ( )
3. Neutral ( )
4. Disagree ( )
5. Strongly Disagree ( )

Here we see when comparing values "4" and "5," for instance, that these numerals not only represent qualitatively different responses but quantitatively different ones as well. Admittedly, then quantitative characteristics are quite rudimentary but this transitivity[1] is important when it comes to our statistical analysis options. A final point to be noted about ordinal variables is that do they not tell that there is a fixed distance between the value. For instance, in the example above we do not know whether the distance or difference between "strongly agree" and "agree" is the same as the difference between "disagree" and "strongly disagree." This occurs because we have no information about the unit differences.

The third level of measurement has all the characteristics of the ordinal level plus the attribute that there are equal intervals or distances between the adjacent values of the variable. The existence of this fixed unit of measurement has led to these types of variables being labelled "interval." When numerals are assigned to the values of an interval scale then the arithmetic opera-

[1] Transitivity is a feature of an ordinal system. Briefly this property is expressed by the rule, "if A is greater than B, and B is greater than C, then A is greater than C."

lowest possible level of measurement. Because levels of measurement place important constraints on the kinds of statistical analysis that can be done, it's a good idea to start with as wide a range of analysis options as are possible, which means with measures of the highest possible level. If, after the data are collected, one chooses to do analysis for which higher levels of measurement are appropriate, it is always possible to create these from lower order measures. For example, if the data on the variable "age" are ratio (that is, if respondents' ages are in years), one can always transform these ages into an ordinal variable by creating categories like:

0 - 5 years = 1
6 - 35 years = 2
36 - 45 years = 3
46 + = 4

However, note that the reverse process is not possible; that is, given age data in the above ordinals categories one is not able to recreate the exact age of particular people. So, to avoid disappointment, keep all options open and measure at the lowest measurement level possible simply because it enhances the power to evaluate causal relationships.

1.5.3 Some Comments on Causal Thinking

Social Science, like natural science, is ultimately interesting in the twin goals of "explanation" and "prediction." Moreover, both explanation and prediction of events rest on the ability to establish the "causes" of these occurrences. For example, if at a cocktail party it is observed that someone drops a glass and falls unconscious on the floor, a reasonable question may be, "Why did this behaviour occur?" Asking this question would be requesting an "explanation" for the event and to satisfy the curiosity to search for the causes or able (where it is then an independent variable). Appreciating this, if the hypothesis is that "poverty causes crime" then poverty is an independent variable, but it becomes a dependent variable if the research question asks "What are the determinants of poverty?" Clearly, theory has a vital role here in setting up causal relationships.

The other point about the identification of independent and dependent arises from the fact that a single dependent variable often has more causes than the single independent variable suggested in the hypothesis. But since hypotheses are commonly interested in exploring the effects of one independent variable at a time, while at the same time taking the effects of other possible causes into account, researchers have decided to name these alternative independent variables involved in a relationship by a separate name, that of "control" or "test." As shall be explored more fully shortly, the variables get their name because they establish the <u>conditions</u> under which the hypothesized relationship is being tested. For example, let's say someone was having trouble with a car and wished to test the relationships between turning the ignition key (the independent variable) and the engine starting (the dependent variable). Now, when this relationship is given a single test, no matter what the outcome, one cannot safely conclude that turning the key always has the same effect on the engine. Clearly, one has only explored this relationship under <u>some</u> conditions, and not others. That is, when "test" or "control" variables (like battery charge, temperature) have changed, perhaps the relationship found cannot be duplicated. Sometimes mechanical problems never reappear at the garage because the environment has altered.

Now that the terms researchers use when they speak about causal relationships between variables have been explained, attention can be turned to what <u>criteria</u>

the case, associations between the independent and dependent variable must be examined more closely to determine the effects of control or test variables on the relationship. How this is done statistically will be discussed later, but let's consider an example that illustrates these ideas.

Statistical analysis can demonstrate that there is a high correlation or association between the following two variables: (1) the number of fire engines at a fire, and (2) the amount of fire damage done. Now, just because this is a fact, it should not be concluded that there is a causal connection between these variables. Discovery of a fire usually prompts many to call the fire department, under the assumption that this will reduce the damage. There are many other examples that demonstrate the idea that we cannot treat mere "associations" as causal connections.

So far, the terminology researchers use when they discuss causal relationships and the criteria they use to demonstrate the same has been discussed. Now something should be said about the <u>context</u> in which the search for causal relationships is conducted. The main point to be made is that scientists usually search for causes in what can be labelled a "sparse causal web." This notion conveys the idea that the theoretical framework that any researcher uses in search for causes is limited. A sparse causal web assumes that the causes of any phenomenon are relatively few in number and independent from one another. For example, an auto mechanic never asks about the owner's astrological sign on the assumption that this variable is irrelevant to the system. So too, with any scientific explanation: it looks for the minimum number of relevant causes.

This use of sparse causal webs has proven very useful in science. An attempt to take every variable into account would soon render any coherent explanation impossible; there is always one more factor to consider. So scientific research disciplines), those sponsoring applied studies should take the results with a grain of salt. This does not mean, of course, that such findings are "useless" since it is always better to have some information than to be ignorant when making decisions. It does mean, however, that consumers should expect the findings of studies they commission or undertake to be of local practical utility since, in most cases, they have studied only one slice of a complex phenomenon. Thus any survey research must also include more general reviews of findings in their area, especially if the results are surprising. Social research studies often do have one very practical advantage that is underrated. As shown below, social research is probabilistic in nature, which means that it operates under the principles of inductive logic. It is in the nature of such logic that universal assertions can be refuted by a single negative case. Interestingly, many people in the area of public policy initiate and operate programs on the assumption that their approach or method be "the answer" to the issue of concern. Under such conditions, a single well designed research study demonstrating that this was not the answer in the case under consideration is sufficient to throw serious doubt on the universal assertion. This is all to say that even if social science research often cannot tell practitioners what the solution necessarily is, it can often tell what does not always work. Unfortunately such conclusions are often perceived as "destructive" (as opposed to constructive) criticism. This is clearly an unfair criticism, since a dampening of the zeal with which the latest fashionable "solutions" in many areas of public policy and administration is often well justified. More importantly, a consultant may "tune" the results to the preconceptions of the client. Partly this is to secure further contract work, but also is the outcome of a natural bias to find positive and significant results. It is difficult to pay $50,000 for a study which does not include public opinion polls and census. For instance, when we read in the newspaper that the most recent gallup poll shows 45 percent of Canadians favouring some politician's view on some issue, we are more informed than before we read the article. Note that we are not told "why" Canadians hold this view, which is what makes such questions simply descriptive. Nonetheless, such information is useful for it is essential to have a careful documentation of the facts of a matter before more ambitious exercises such as explanation are undertaken.

"Explanatory studies" comprise the second type of survey research studies. Such studies move beyond descriptive investigations in an attempt to establish causal connections between variables. In order to undertake this sort of investigation researchers are required to have quite specific ideas or hypotheses about what independent variables might be influencing the dependent variable under consideration. As shall be mentioned shortly, this type of investigation presents considerably more obstacles for survey research than it does for other methodologies, like experiments for instance. However, in many instances, surveys present the only feasible approach and can be used to address such issues.

The third general type of survey research studies are called "explanatory" and these, as the name suggests, are useful when relatively little is known about the topic of interest. Under such circumstances broad questions can be asked of members of the relevant group and their thoughts on general issues surrounding the topic collected. From this information it is often quite possible to glean insights into some of the major issues and relationships surrounding the topic. Data can then be used to develop a methodology which would permit a more systematic investigation of the problem.

Let's say we were hired to undertake an investigation of Figure 1 and we chose as our unit of analysis the individual parts labelled A, B, C, and D. Further assume that the variable of interest is "shape." What conclusion might we draw from the inquiry? Simply, that each part has the shape of a triangle. This is fine. However, assume this information is given to someone else and they were asked to draw a conclusion about the shape of the whole Figure 1. What might their answer be? Logically, they may infer that since piece A is a triangle, piece B is a triangle, and so are pieces C and D, then it must be the case that the figure is a triangle - which is clearly incorrect. The reason for the error stems from the fact that we are shifting from one unit of analysis (the parts) to another (the whole) and this often leads to problematic inferences. In science, as in life generally, "the whole is more than the sum of the parts;" a symphony is more than a collection of notes, a painting more than a set of brush strokes, and a mind more than a collection of atoms. So, care must be taken to focus attention on the unit of analysis of interest and confine inference to that unit and no other. An example from the social sciences should solidify this conclusion. Assume we are interested in the relationship between race and crime. The police records divide the city into "neighbourhoods" and that statistic can be obtained on the proportions of Native peoples and non-Native people in each area as well as the crime rate for each jurisdiction. These data may show that as the proportion of Natives in an area increases, so does the rate of crime. This, by itself, is a fair conclusion. What cannot legitimately be concluded from these findings is that the crimes in those areas having a high proportion of Native people were committed by them. Such a conclusion is unscientific since it involves a shift from the unit of analysis "neighbourhood" to that of "individuals." In largely Native areas it may be the non-Natives who spuriousness. With cross-sectional designs it is quite straightforward to establish the just criterion of association, as shall be demonstrated in a later section of this manual. The latter two criteria, however, are considerably more difficult to handle with cross-sectional data.

Since data for cross-sectional studies are collected at one point in time, right away one faces a problem in demonstrating that the independent variable(s) took place earlier in time than the dependent variable. After all, in this design one has no direct means of controlling or manipulating the independent variable to assure it occurs prior to the dependent variable, as in an experiment, for instance. Consequently, only indirect methods of meeting the time-ordering criterion can be used, some which are common. First, sometimes a logical argument can be made to demonstrate the required time-ordering. For example, if the two variables were "child rearing practices" and "job satisfaction" it could be argued that it is logically impossible for present job satisfaction to have been the cause of socialization by one's parents and, thus, that child rearing practices must be the independent variable in this relationship. In many cases, however, it is logically possible for either variable to have occurred just as is the case with the variables "drug addiction" and "high school drop out." Quite possibly if students become drug addicts this so affects their performance that it eventually causes them to drop out of high school. On the other hand, perhaps it is the case that high school dropouts find themselves in social situations that increase the probability of their becoming addicted to drugs. In cases like these some approach other than logicality must be employed to establish time-ordering. A second such possibility involves the use of retrospective questions.

who began university three years previously. Of course, it is quite possible that the existing fourth year students are only composed of those with above average intelligence since the others would have failed out of university during the course of their studies. If this is the case then cohort comparisons in cross-sectional data compromise their validity.

Besides the problems with establishing time-ordering or sequence using data from cross-sectional survey designs, there are also difficulties with meeting the third causal criterion, that of non-spuriousness. In order to establish non-spuriousness one must rule out possible alternative variables that might explain the hypothesized relationship between two variables. In experimental methods, this is quite simple since we can make the comparison groups equivalent on all variables except the independent variable and thus conclude from the data that any observed differences are due to the effect of the independent variable. Such strict controls are not available to survey researchers using cross-sectional designs, so some other technique must be used. This alternative technique involves the use of introducing statistical controls when analyzing relationships. These techniques will be introduced in the analysis section of the text. For now only one point is essential. In order to statistically control for the effects of possibly contaminating variables one must have information from the sample on these variables. This, in turn, requires that such data be collected at the time when the survey is being conducted. To accomplish this, one must have some idea of which variables may affect the hypothesized relationships and at this point the researcher has problems. For if it were known in advance which variables would affect the hypothesis, one would probably change the nature of the hypothesis. Since the hypothesis contains the "best-guesses" about anticipated relationships there is no reliable way of "knowing" for what samples are drawn on several occasions. The essential difference between these two designs concerns the focus of the sampling procedure. In trend studies the entire population is sampled at different times, whereas in cohort studies only one subpopulation is tracked. Members of a cohort (or those sharing a common characteristic) constitute the particular subpopulation used and perhaps the most common type of cohorts use age as their focus. For example, in the trend study cited previously, at each of the sampling points a representative sample of Canadians was drawn. In doing a cohort study of the variable, "attitudes toward abortion," we might initially sample those who were thirty years old in 1950. For the next sample, to be taken in 1960, we would sample from the <u>same</u> subpopulation who, at this time, would be <u>forty</u> years old. And then, in 1970, we would draw a random sample of those fifty years old and over about their attitudes toward abortion. Similarly we would complete the study in 1980 by sampling those who were sixty years of age. Note that cohort studies, like trend studies, provide us with a descriptive summary of how some characteristic is changing over time but the description is based on people sharing a particular characteristic, not the general population.

Both trend and cohort studies share an advantage over the type of descriptive information provided in cross-sectional designs; namely, their ability to analyze change over time. The fundamental limitation of trend and cohort studies is that they only provide the aggregate picture of how groups are changing; they cannot provide specific insights into the nature of change among specific individuals. It is for the purpose of studying individual change that longitudinal <u>panel</u> designs have been developed. The reason these are called "panel" studies comes from the fact that data are collected from the same people on several occasions. Through collection of data from panel members who are, for example, survey researcher must be aware of the strengths and limitations of cross-section and longitudinal studies.

The next step is usually the translation of the hypothesis into a questionnaire. In the next chapter, the three major questionnaire designs - mailout, telephone and in-person interview are discussed in depth.

Chapter 2

QUESTIONNAIRE DESIGN

2.1 INTRODUCTION

Once the decision to conduct a survey has been made, attention naturally turns to several practical questions. How is the sample to be drawn? When is the best time to conduct the survey? How fast can it be completed? How much will it cost? However, one of the most important concerns centres around the actual questionnaire itself. For example, how long should the questionnaire be? Should it be mailout, or should it be completed by trained interviewers meeting face-to-face with the respondents? Are answers given over the phone permissible? Does it make any difference who answers the questions? Will it make a difference if the order of the questions is changed? And so on!

The questionnaire design is critical. Indeed, it is the questionnaire, more than any other piece of information perhaps, that most accurately captures the objectives and rationale of the survey. The questionnaire is the document most frequently reqested by others after the survey is over; it is the most carefully studied and useful piece of information for any subsequent appraisal of the survey results as well. While other elements such as the sampling design, etc., are very important, the questionnaire design can make or break the survey. An improperly worded or formatted questionnaire may result in so few useful responses that interpretation becomes impossible. In sum, the questionnaire design is a critical component of the entire survey process.

Special issues peculiar to mailout surveys and telephone surveys are mentioned next. Finally, no matter how much care and thought is put into the task of developing a questionnaire, there is no satisfaction if the questionnaire remains unanswered. Accordingly, methods and techniques "to enhance response;" that is, to encourage as many people as possible to answer as many of the asked questions as possible, will be mentioned. Nevertheless, it is rare that response to a questionnaire is totally complete. There will be some questions that people will refuse to answer, and there will always be some people who will refuse to be interviewed at all. Some thought must be given then to the matter of how the results of the survey may be interpreted under these circumstances.

2.2 FACE-TO-FACE VERSUS MAILOUT VERSUS PHONE INTERVIEWING

Which is the best survey method to obtain the most reliable and most complete information? The question is obviously "loaded." A researcher wishing to conduct a survey of the 1960 graduating class in medicine probably has no choice - the mail questionnaire. A government official wishing to guage the public's reaction to last night's announcement of wage and price controls by the Prime Minister and have the results tabulated by next week's Cabinet meeting also has no choice - the phone interview. And finally, there is often no substitute for in-person interviews for special population groups, for example small children.

But which method should be chosen when there is a choice? The answer is still: "It depends!" However, there are a number of attributes of each survey method that can be considered - the likely response rate, the overall cost, the logistics, the sampling constraints, etc. This section will consider the relative merits of the mail, telephone, and face-to-face survey methods from the special perspective of questionnaire design and content issues. The discussion

2.2.2 Length of Questionnaire

It is usual to hear the general advice to "keep it short;" the major argument is that lengthy questionnaires create fatigue among respondents, result in poorer data quality, and lead to higher rates of refusal and non-response. Length is usually not a serious problem for face-to-face interviews; marathon sessions of four to five hours are not impossible. Many an interview supervisor has heard the lament of interviewers who insisted that only one or two interviews could be accomplished in the time allotted by the supervisor for three or four interviews. Generally, face-to-face interviews of one to two hours are fairly routine to implement. On the other hand, while high response rates (averaging 75 percent) have frequently been achieved with lengthy mail questionnaires, an overall examination of various mail surveys tentatively concluded that about "11 pages, or 125 (question) items, represent plateaus beyond which response rate reductions can be expected" (Dillman, 55). Finally, the length of telephone interviews does not appear to be a problem. Hour-long interviews have presented no difficulty in some cases, and interviews of 20 minutes duration are routinely administered. Perhaps because of the "social" courtesies that are often observed on the phone, the evidence suggests that once people are on the phone, they will rarely terminate an interview because of its length.

2.2.3 Types of Questions

Because each survey method uses a different mode of communication, there are differences in the type of questions that can be asked. Questions differ with respect to their complexity, their nature (open-ended or not), their purpose (screening or skipping), their sequence, their tedium, their potential for non-response, and their threatening stance, to name but a few characteristics. Each survey method handles these with varying degrees of success.

cannot be controlled at all since the respondent may read the entire questionnaire before answering. Furthermore, respondents may also go back and change their answers to earlier questions. Tedious or boring questions are a particular problem for mail surveys because of the extra motivation required when compared to the in-person or phone interview. Questions may also be skipped more frequently since there is no interviewer (in person or phone) present. However, "threatening" questions are sometimes better handled by mail questionnaires than face-to-face interviewers.

The phone interview shares many of the advantages of face-to-face surveys. Questions may be complex, open or highly detailed. The sequence can be controlled and explanations, probes and the like are possible. On the other hand, visual aids of any sort are impossible, and questions in which the respondents are asked to perform rankings or select an item from a long list of categories especially tax the respondent's memory and patience. Further, it is generally more difficult to communicate detail over the phone, respondents may be too embarrassed or simply reluctant to ask interviewers to repeat questions.

Finally, all three survey methods have been used with various degrees of success for surveys of special populations. Broad generalizations are impossible, and depend upon the group in question. In-person interviews may be most effective with some ethnic groups, provided that interviewers can be found to overcome the language problem. Phone interviews may be the only possible avenue to survey populations which are highly mobile (e.g., migrant workers), and self-administered questionnaires may be required strictly on logistic considerations (e.g., large numbers of students).

racy and validity. If it is remembered that asking and answering a set of questions is an act of social interchange, it is clear that specific answers to questions may be subtly influenced by survey method. Questionnaires may be subverted by the respondent, the interviewer or others.

Respondents may offer what they perceive to be socially desirable answers, in other words, what they think is the "right" answer or the answer they think the researcher "wants." The face-to-face interview is most likely to elicit the socially desirable biased response. The phone interview appears to decrease the likelihood of this happening, but often to a negligible extent (Dillman, 62). Generally, people are most honest in mail surveys, least honest in face-to-face interviews, and intermediate in the phone interview. In summing, face-to-face interviews have the highest probability of producing socially-desirable answers. It is especially important to bear this in mind when asking intimate or personal questions.

Interviewers also create distortions. Some of it is unintentional when poorly-trained interviewers freely interpret the meaning of the questions. Interviewers may even deliberately fabricate interviews. Although proper training and quality-check monitoring procedures will help, the potential for interviewers to produce inaccurate answers is there nonetheless. A centralized and closely monitored phone interview team is less likely to produce interviewer-induced distortion and subversion. Uniformity in the way a question is asked can only be achieved in the strict sense by a mailout questionnaire. Here again, there is a price to pay for the absence of the interviewer. The self-administered questionnaire may contain useless answers when questions are left to the respondent alone to interpret.

from which to draw. Transportation also poses a problem. Telephone interviewing is relatively simpler, where the skill requirements are generally less and central supervision allows easy consultation with researchers and interviewing schedules are usually easier to accommodate. Personnel requirements are even less onerous for a mail questionnaire.

As a general rule, telephone surveys are the fastest to complete. Mail surveys must typically follow a definite schedule for mail, re-contacts, and other follow-up procedures. Only for very large surveys would a mail questionnaire compare favourably with a telephone survey in terms of speed. Face-to-face interviews are typically time-consuming and cannot be rushed. Weather, geographical dispersion, the nature of the group surveyed, availability of interviewers, etc. - all are factors determining the speed with which a face-to-face survey can be completed.

It is definitely more expensive to conduct face-to-face surveys, but the costs compare favourably to other methods if there is minimal geographical dispersion. On the other hand, geographical dispersion is virtually irrelevant for mail surveys since postage and material costs are identical for a province-wide or nation-wide effort. Telephone interview costs obviously will differ whether the sample is "local" or "long-distance."

2.2.7 Summary and Conclusion

This presentation of the relative merits of face-to-face, mail, and telephone survey methods has been by no means exhaustive. Nevertheless, it should be clear that no single survey method is best under all circumstances. The exasperating answer to the question of which method is best must remain: "It depends, and that's definite."

TABLE 1

RATING OF SURVEY METHODS FOR SELECTED PERFORMANCE CHARACTERISTICS

Performance Characteristic	Method		
	Face-to-Face	Mail	Telephone
1. Allowable Length of Questionnaire	High	Medium	Medium
2. Type of Question			
(a) Allowable Complexity	High	Medium	Low
(b) Success with Open-Ended Questions	High	Low	High
(c) Success with Screens	High	Medium	High
(d) Success with Skips	High	Medium	High
(e) Success with Sequence	High	Low	High
(f) Success with Tedium	High	Low	Medium
3. Success in Avoiding Item Non-Response	High	Medium	High
4. Sensitivity to Construction Procedure (Format)	Low	High	Medium
5. Success with Threatening Questions	Low	High	Medium
6. Success with Special Population Groups	Indeterminate	Indeterminate	Indeterminate
	(depends on the population)		
7. Accuracy and Validity			
(a) Social Desirability Bias	Low	High	Medium
(b) Avoiding Interviewer Distortion and Subversion	Low	High	Medium
(c) Avoiding Contamination	Medium	Medium	High
(d) Allowing Consultation	Medium	Medium	Low
8. Administrative Considerations			
(a) Difficulty of Personnel Requirements	High	Low	Medium
(b) Speed of Implementation	Low	Low	High
(c) Potential for Low Costs	Low	High	Medium

3. A MODERATE PROBLEM

4. A SERIOUS PROBLEM

Some people might respond "a serious problem" because they believe illegal abortions are numerous and produce medical problems; others might respond that it is "a serious problem" because they feel illegal abortions are morally wrong - an attitude rather than a belief.

Questions that seek to capture information about behaviour or attributes are relatively uncomplicated. For example, the question: "Have you ever had an abortion?" is directed towards the respondent's past behaviour. However, questions concerning behaviour may also concern people's present behaviour (e.g., "Are you currently working at a job for pay?") or future behaviour (e.g., "Will you be leaving the province of Manitoba permanently during the next six months?").

Attributes are also commonly referred to as personal and demographic characteristics. Questions which ask the respondent's age, sex, marital status, race, income and the like are frequently encountered in surveys; their usual purpose is to relate how the other information gathered differs for people with different attributes.

2.3.2 Question Structure

A major criterion for distinguishing between question structure is whether the nature of the response is <u>open</u> or <u>closed</u>. An open question provides no fixed list of answers from which the respondent chooses; instead, the respondent "creates" an answer. In contrast, a closed question provides a list of answer

- Differences in interpretation of question go undetected
- Variation in answers are artificially eliminated
- More likelihood of transcription error

The advantages of open questions are:

- May be used when all possible answer categories are unknown
- Allows respondent to answer in detail, to clarify, to qualify
- May be used when there are too many potential categories
- Some issues are too complex to condense into fixed categories
- Allows creative, natural self-expression for respondents

The disadvantages of open questions are:

- Yields collection of worthless and irrelevant information
- Data not standardized from person to person
- Coding is difficult, often subjective, leads to error
- Requires superior writing and communication skills
- May be too generally worded to convey meaning
- Requires greater effort and time of respondent; possibly higher refusal rate
- Questionnaire may require more paper; looks longer and more formidable.

The above discussion has focused primarily on the distinction between open and closed questions and their advantages and disadvantages. There are also hybrid type questions. These are some closed questions in which the answer categories are "unordered." For example, the question: "Which best describes the kind of house in which you live?" and the answer categories provided are: "single family dwelling," "duplex or triplex," "apartment" and "mobile home." Similarly, some questions may be only "partially closed." For example:

or "scale." As mentioned in the previous section it is usual to distinguish the following levels of <u>measurements</u>:

<u>nominal variable</u> (discrete non-numerical categories such as male or female)

<u>ordinal variable</u> (rank-ordered categories such as most favourable, neutral, least favourable, etc.)

<u>interval variable</u> (ordered categories with equal intervals between ranks, such as Centigrade temperature)

<u>ratio variable</u> (interval variable with a fixed zero point such as income, age, etc.)

For nominal questions with factual answers it is usual to list all possible answers and provide a blank, a box to be checked, or a number to circle. For example:

Indicate your sex by circling the appropriate number.

1. MALE

2. FEMALE

It is also common to place response categories for other types of variables in a similar format. For example:

How often to you read McLean's?

1. REGULARLY

2. OCCASIONALLY

3. HARDLY EVER

4. NEVER

A. Strongly Agree / Agree / Neutral / Disagree / Strongly Disagree

B. Excellent / Good / All Right / Poor / Bad

C. Often / Sometimes / Almost Never

D. Very Important / Important / Somewhat Important / Not Important

Sometimes, a researcher will wish to obtain a fairly fine degree of detail, but find it impossible to construct a series of words which reflect the gradation desired. This desire for a "continum" has often led some researchers to provide a "thermometer" response scale with labels only at the extremes. Consider the following example:

> To what extent should the Provinces have exclusive control over off-shore resources rights in Canada?

PROVINCES		CANADA
Should have Exclusive Control		Should have Exclusive Control

/ / / / / / / / / /

0 1 2 3 4 5 6 7 8 9

(CIRCLE APPROPRIATE NUMBER)

For a more complete discussion of the theory and practice of measurement in social surveys, see Anderson, Basilevsky and Hum (1982).

2.3.4 Wording of Questions

One of the most time-consuming and difficult tasks in conducting a survey involves the wording of questions. The wrong choice of words can create numerous problems. The question may not be understood; it may be misinterpreted; it may class, or Lower class?" is ambiguous. So too are questions concerning race or skin colour. In addition, words which have slang connotations should be avoided; for example, "bad" has negative connotations in ordinary usage but can have either positive or negative connotations in Black slang.

3. <u>Is the Level of Wording Appropriate</u>?

It is important to write a question that means the same thing to everyone. For this reason, substitute words for needlessly complex expressions. For example, "honest" can be used instead of "candid," "brave" instead of "courageous," etc. It is not always the case that shorter words are preferable, however. It is quite proper to use "annexation" rather than "an addition" when interviewing city planners, or "pharmaceutical companies" rather than "companies that sell medicines" when interviewing medical doctors.

4. <u>Is the Question Too Abstract</u>?

Factual questions usually are quite concrete. However, questions concerning feelings or opinions are especially difficult. Respondents might find it difficult to answer questions about their "satistifaction," "annoyance" or "happiness." For example, "How happy are you?" Very happy, moderately happy, unhappy? is too abstract.

5. <u>Is the Question Too Leading</u>?

Leading questions increase the probability of a biased response. Questions should be asked in as neutral fashion as possible. For example, one should ask: "Do you smoke?" rather than "You don't smoke, do you?" or "National Health and Welfare tells us that smoking is harmful, do you agree?".

How many books did you read last year?

1. NONE

2. 1 - 10

3. 11 - 25

4. 26 OR MORE

8. Is the Question Biased?

A biased question is one which is worded so as to encourage respondents to answer in a specific fashion. Leading questions can often lead to bias. Bias can also result from creating a behavioural expectation; for example, "The majority of Canadians believe that the RCMP should not pay convicted murderers for information concerning the location of the victims' bodies. Do you agree?" Bias can also result from unequal comparison; for example, "Who do you feel is most responsible for the high cost of mortgage rates in Canda? FARMERS, LABOURERS, CHARTERED BANKS. Finally, bias may be caused by using an unbalanced set of categories. For example:

Currently Canada transfers to the Province about 5 billion dollars a year for health and education programs. Do you feel this amount should be:

1. INCREASED SLIGHTLY

2. STAY THE SAME

3. DECREASE A LITTLE

4. DECREASE SOMEWHAT

5. DECREASE A GREAT DEAL

"Unwed mothers on welfare are parasites on society."

1. AGREE
2. DISAGREE

Note that the above revision employed a series of questions to reduce the objectionable nature of a question. Brevity is not always a virtue. Even so it is wise to consider very carefully all questions which may be objectionable.

10. <u>Is the Question Too Demanding</u>?

Questions that try the patience and mental agility of the respondent are unlikely to be successful. For example, asking the respondent to rank the 10 provinces in order of their commitment to preserve the environment is too demanding. Similarly, asking a respondent: "What percentage of your monthly income is spent on rent for your house?" requires a great deal of thought and calculation by the respondent. Again, switching the question format is possible.

How much is your average monthly income?

____________________Dollars

How much do you pay each month for rent?

____________________Dollars

11. <u>Does the Question Require Two Answers</u>?

Frequently, a question is worded in such a way that the answer categories are not mutually exclusive. For example, when numerical categories overlap, such as "18 - 25 years of age," "25 - 35 years of age," etc. At

The Crow's Nest Rate is an agreement whereby .... The Premier of Manitoba has recently taken a stand on this issue. Were you aware that he had taken a stand on the Crow's Nest Rate?

1. NO (GO TO QUESTION ___)

2. YES (IF YES, Please describe in you own words what you think his position to be).

____

Do you tend to agree or disagree with his stand?

1. AGREE

2. DISAGREE

13. Does the Question Provide an Appropriate Time Referent?

Questions that contain a time referent may introduce variations in answers, especially if the survey is conducted over a lengthy period of time. Similarly, key dates may be useful in wording questions. For example, "How many times did you attend a movie during 1981?" or "On the average, how many times a year do you attend movies?" is preferable to "How many times have you attended movies this year?"

Questions which refer to past periods are also troublesome. It is especially helpful if the reference period can be related to a significant event. Following are some examples.

Were you a resident of Manitoba in 1963? That was the year President Kennedy was killed.

new policy option (say, a tax credit). Perhaps the response category format should then be extended, or a user of the data may explain some subtle point which reveals the researcher's ignorance of the situation (say, a fine point of the law). Finally, the questionnaire should be administered to a number of representative respondents. Not only would field procedures be worked out, but also respondents can indicate which questions are obscure, which are hard to interpret, and the like.

2.4 MAILOUT SURVEYS: SELECTED ISSUES

Special care must be taken in the construction and administration of a mail questionnaire. This is because the questionnaire must "stand alone." Unlike the in-person or telephone survey, there is no interviewer present. Consequently, design deficiencies in the questionnaire cannot be ameliorated by the various aids and support which an interviewer can give. This fact more than another - the absence of an interviewer - creates a number of special problems for mail questionnaires. This section indicates a few selected issues having to do with mailout surveys.

2.4.1 The Importance of Format and Lay-Out

Because the mail questionnaire must be its own advocate, it is important that the format of the questionnaire be attractive and that the lay-out of the questions be easy to read, pleasing to the eye, etc. This does not mean that a graphic artist must be a member of the survey team, but clearly a sense of design must be borne in mind. Whether the questionnaire is to be printed or reproduced by other means will depend, of course, on sample size and cost considerations. Nonetheless, it is helpful to make the questionnaire appear as professional as possible. Some suggestions include:

3. Ask screen questions; that is, those which determine whether the respondent is eligible to answer, at the beginning.
4. Questions covering the same topic should be grouped together. Not only will this be less demanding for the respondent, it may also conserve space and make the formatting of the response categories easier.
5. Be careful to provide clear instructions to questions involving rankings (high to low or low to high), items in a grid, etc. Do not assume the respondent will "figure it out," to check or circle only one item, or all items that apply, etc.
6. Use transitions to establish continuity of questions or to introduce a new topic. For example, "Next, we would like to ask you several questions about how you spend your leisure time."
7. Place personal questions at the end. This will minimize non-response since many respondents "do not see" why questions about their "income," etc., have anything to do with the subject matter of the survey (e.g., their annoyance with aircraft noise, etc.). A transition is usually helpful: "Finally, we would like to ask a few questions about yourself for statistical purposes (to help us interpret the survey results), etc."

2.5 TELEPHONE SURVEYS: SELECTED ISSUES

It has been believed that telephone interviews are "inferior," must be "short and quick," and ask only "simple" questions. The "bad press" given to telephone interviewing is unjustified in light of new developments, empirical evidence and experimental studies (See Dillman, Quinn, Gutek and Walsh, for example). More and more attention is now being paid to telephone interviewing for the following reasons: (1) response rates are often low for face-to-face interviews, (2) tion. I would like to tell you about this proposal announced in the Budget and find out how you feel about it. Here is what it would do.

First, many items exempt from income tax would now become taxable. Also, the income tax rates would be lowered at the higher end.

Second, the federal government wants to establish conditions as to how the money given to the Provinces would be spent for education purposes. If the provinces do not reach agreement in two years, the Federal government would freeze the amounts given to the provinces at that level.

So, what has been proposed is to make more income subject to tax but lower the tax rate, and to impose conditions on the province as to how they spend their money for education purposes.

How do you feel about these proposals? Let's consider the move to lower tax rates but eliminate some tax deductions. Are you strongly opposed, somewhat opposed, uncertain, somewhat in favour, or strongly in favour?

STRONGLY OPPOSED	1
SOMEWHAT OPPOSED	2
UNCERTAIN	3
SOMEWHAT IN FAVOUR	4
STRONGLY IN FAVOUR	5

4. Separate complex questions into simpler ones; the above question is a good example. Also, note that the numbers identifying the response categories have been placed to the right to facilitate coding by the interviewer. Similarly, instructions to the respondent (such as CIRCLE ONE) are omitted.

5. Avoid excessive response categories. While it may be possible in a mail questionnaire to include seven or eight response categories, this list would be too long for a telephone interview. A maximum is probably five.

NOT ENOUGH 1

ABOUT RIGHT 2

TOO MANY 3

Of the channels in Winnipeg, is there one that you tend to watch more often?

NO 1

YES 2

Which channel is that? _______

Discussion is now going on about Pay TV. This would be a special channel with no commerical messages. However, subscribers would have to pay a monthly fee in order to receive this channel. In general, what do you think of such a channel for Winnipeg?

2.6 SUMMARY

This chapter has presented the basic concepts of questionnaire design. As in any complex task, practice does improve performance. Questionnaire design may be one of the few things a committee does well (the other might be "passing the buck"). Next, we turn to sampling and defining the particular mathematical (probability) foundations of the survey.

Chapter 3

SAMPLING: THE KEY TO SUCCESS

3.1 INTRODUCTION

The types of information that may be obtained or viewed as important are many and varied. General impressions and feelings may be appropriate at some levels of our decision-making processes and may be extremely useful in providing guidance as to appropriate action. In particular, most often what is required is information concerning a total, an average, or a proportion. Wildlife conservation officers require information about the total number of deer harvested annually by licensed hunters. A social agency may require information about the average family income in an area; and an advertising agency may want to know what proportion of the people of a given age listen to a particular radio station. It is the purpose of this section to examine the way in which information about such items, as indicated above, can be obtained in an efficient and economic way.

3.2 ELEMENTS OF SAMPLING

3.2.1 The Population or Universe

For our purposes, a population or universe is considered to be the totality of all items about which we wish to obtain some information. The population will not necessarily refer to people but may, in fact, be a collection of trees, all single family dwellings in a city, business enterprises, fish in a lake, to name but a few.

3.2.2 Advantages of Sampling

1. Cost

A sampling procedure will include unique cost considerations involving the planning of the survey, and selection of the sample units, costs which are not encountered in a complete enumeration survey. However, given a fixed per unit cost for extracting the information from a unit, the total cost of the sample survey should be less than the cost for a complete enumeration by virtue of the fact that fewer units are examined under sampling.

2. Speed

The sampling procedure involves fewer units which should result in the completion of the collection and analysis of the data in a shorter period of time.

3. Accuracy

With the requirement of examining fewer units, there is the potential for greater control over the collection of the information, the administrative details are less onerous, and the potential for human error is reduced.

4. Scope

With the increased time and effort that can be spent in collecting information from the sampled units, there is a greater potential for obtaining extensive and detailed information not possible under complete enumeration conditions.

then there are many situations where a sample will provide more information about a population than a census. That is, it makes sense to exchange somewhat higher sampling errors for greater information. This may be an obvious point, but it is frequently ignored or misunderstood.

The procedures to be outlined in this section are designed to enable the research worker to control, to reduce, and most importantly to measure the sampling error. With this control, and with the potential for greater accuracy, reduced cost, and increased scope has come the widespread acceptance of sampling as an important tool for the gathering of information.

3.3 TYPES OF SAMPLING

3.3.1 Probability Sampling

Probability sampling refers to a procedure whereby the units chosen from the population to form the sample are selected by a chance mechanism. There are known probabilities of selection assigned to each and every unit in the population, and a process of selection is devised which guarantees that only the selection probabilities are used in the selection process. The simplest of all such selection procedures is a simple random sample in which all members of the population have an equal chance of being selected.

3.3.2 Non-Probability Sampling

Non-probability sampling does not have the feature of selection of units according to known probabilities. There are many examples of such non-probability sampling schemes, and their popularity is widespread. Falling into this category of non-probability samples are the following:

as <u>expert sampling</u> as well, to reflect the fact that frequently someone with specialized knowledge of the population of interest is used to determine which units in the population would provide the "best" or "most accurate" information.

3.3.3 <u>Probability Sampling Versus Non-Probability Sampling</u>

Non-probability sampling procedures such as judgment sampling may provide useful information about the population. The prime difficulty with such sampling, however, is that there is no objective way of measuring the precision of the results. The results of a particular survey often are not necessarily reproducible since a second expert may disagree with the first and suggest a quite different sample of units producing very different estimates.

In addition, there is a strong possibility that biases may be introduced by the researcher. With volunteer sampling one must always question why an individual might volunteer. It is well known, for example, that successful deer hunters are more likely to volunteer the information of their success, while unsuccessful hunters are much less likely to divulge their failure. The results of such a volunteer survey are thus known to over-estimate the success rate resulting in a very high estimate of total hunter harvest.

A probability sample however, has the desired feature that all units selected were done so by a chance mechanism alone. The known selection probabilities enable us to use the mathematical theories of probability to assess, objectively, the sampling variability. The estimates obtained from the units will have properties that either are known, or are capable of being determined.

While there are situations in which it may be acceptable to use a non-probability sample, for instance in preliminary or exploratory work, or in surveys

3.4 THE STRUCTURE OF A SAMPLE SURVEY

A sample survey consists of a number of well defined steps or procedures:

1. A carefully outlined set of objectives for the survey;
2. A definition of the population to be surveyed or sampled; it is often the case that in order to achieve the objectives established for the survey there is a defined population known as the target population, from which the data are to be obtained. In practice, however, the population from which the sample is, in fact, drawn, known as the sampled population, may not coincide exactly with the target population. It is important to note that if this is the situation at hand, any conclusion drawn from the sample will refer to the sampled population alone. Care must thus be taken to ensure that differences between the target population and sampled population are not extreme.
3. Decisions made concerning the data to be obtained. The instrument which will be used to extract the information from the sampled units will depend upon the nature of the data to be obtained. The amount of information must be determined, and the precision with which population parameters are to be estimated should be specified.
4. The defined population is partitioned into non-overlapping, exhaustive parts called sampling units, or simply, units. Such units may consist of individual items or people, or may consist of groups of items, such as city blocks of houses. A listing of all such sampling units, called the sampling frame, is then obtained. It is from this frame that the sample is to be selected, so that it becomes one of the very crucial and key requirements of the survey design.

size together with a preassigned probability of its selection, a mechanism by which a particular sample is selected using the assigned probabilities, and a specification of the nature of the estimation procedure to be used following the obtaining of the appropriate data from the selected sample.

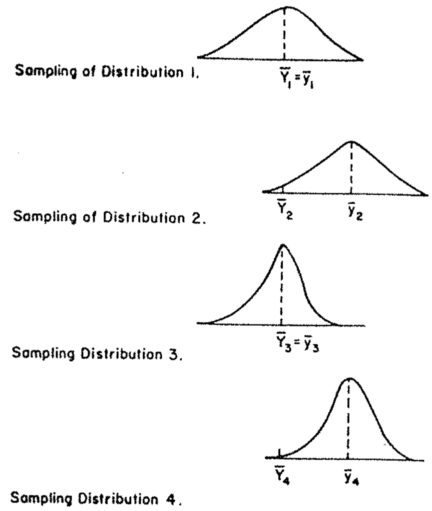
5. The administration of the survey. This will consist of such items as a pre-test of the instrument for collecting the information; the actual selection of the sample; the training of interviewers; the supervision of the collection of the data; the editing, and the preparation of the data for analysis.
6. The analysis of the data. The data are analyzed in a manner appropriate to the design and estimates of the population parameters to be obtained. The analysis should provide for some measure of the error in the estimate to be reported.
7. A careful review of the entire procedure. This can provide important information for future surveys on a similar or related population.

3.5 SOME ADDITIONAL TECHNICAL MATTERS

The data from the sample obtained are subjected to analysis, and, in particular, estimates of the population parameters are obtained. In order to assess the effectiveness of our sample in providing information about the parameters, it is necessary to consider some properties and to examine the behaviour of the estimates.

To assess an estimator of a population parameter, consider a population of households with the parameter of interest being the average annual income of all households. On the basis of a random sample, the sample average income per hou-

Figure 1



These four sampling distributions illustrate some very important features. We just note the location, as indicated on each figure, of the true population average, $\bar{Y}$. In relation to this, we note that the centre of the sampling distributions 1 and 3 correspond to the location of $\bar{Y}$, whereas in sampling distributions 2 and 4, the centre of the distribution, designated by $\bar{y}$, differs from $\bar{Y}$. The difference $\bar{y} - \bar{Y}$ is called the bias of the estimator.

Also, if we consider the degree to which the sampling distributions are concentrated around their centre, sampling distributions 1 and 2 have a certain amount of spread which, in contrast to sampling distributions 3 and 4, is much larger. The measurement of this spread, or variation, around the centre of the sampling distribution is called the variance of the sampling distribution, and the square root of the variance is called the standard deviation, or the standard error of the sampling distribution of the estimator.

Given a choice, it is desirable that our estimator and sample design be such that the bias in the estimator be small, preferably zero, and the standard error of the sampling distribution also be small. Of these four illustrations, the fidence that the interval encloses the unknown value of the parameter. This interval, a <u>confidence interval</u>, may be constructed using the properties of the normal distribution, the standard error obtained from the sample, and the sample estimate of the parameter.

Example,

> In particular, were we to obtain an estimate of the average household income from our sample, which had the value of say \$12,100.00 with a standard error estimated at \$800.00, we can construct a 95 percent confidence interval estimate. The form of the interval would be:
>
> $$\$12,100 - 1.96\ (\$800) \leq \bar{Y} \leq \$12,100 + 1.96\ (\$800)$$
>
> which produced the result:
>
> $$\$10,532 \leq \bar{Y} \leq \$13,668$$
>
> where the number 1.96 is obtained from properties of the normal distribution. Note that we infer this from the Central Limit Theorem and properties of the <u>sample</u>, namely the average income and standard error (\$14,100 and \$800 respectively).

The appropriate interpretation for an interval statement of the above type is that in repeated sampling using this sample design and sample size, it is expected that 95 percent of all intervals determined in this way will enclose the true value of the parameter. Thus, we state that we are 95 percent confident that this particular interval (\$10,532, \$13,668) covers the unknown parameter value. Note, however, that the interval may <u>not</u> enclose the true value, and we have absolutely <u>no way</u> of knowing if this is the case. However, if 95 percent of all such intervals do, our confidence is high (i.e., 95 percent) that the particular interval we produced is one of the "good" intervals.

As will be illustrated later, the concept of a confidence interval, and, moreover, the length of a confidence interval, give guidance in answering the very important and difficult question on the correct sample size.

There are, in fact, two situations which must be considered; the first allows for any unit in the population to appear any repeated number of times in a sample. Such a simple random sample is called a Simple Random Sample With Replacement (SRSWR). The second restricts a unit so that it can be selected, at most, once in any possible sample resulting in the sample of size n, consisting of n unique units. This simple random sample is called a Simple Random Sample Without Replacement (SRSWOR).

In practice, a simple random sample is selected unit by unit from the frame, with the selection being done so that each and every unit has the same chance of being selected. To accomplish this, one might assign a unique number from 1 to N to each unit in the frame. Following this assignment, a table of random numbers can be consulted and from the table a set of n numbers selected.

More common is the use of computer procedures which will select a random sample from a list using a pseudo random number generator. Typically such pseudo random number generators will produce non-repetitive sample random numbers between 1 and several million and are more than adequate for use in social surveys.

These selected numbers will correspond to units in the population and the designated units will form the Simple Random Sample. If a unit may be selected more than once we will have selected a SRSWR, however, if repeated units are rejected, the sample will be a SRSWOR.

Generally speaking, the SRSWOR, of a particular size, is preferred over an SRSWR of the same size, as the former provides greater precision in the estimator. The standard error of the estimate will tend to be smaller for the SRSWOR by virtue of the fact that there will be n distinct units, whereas the SRSWR may have fewer distinct units in the sample of size n and provide, therefore, less information about the population parameter.

The population units are, at least theoretically, numbered from 1 to N. To select a sample of a given size, a unit is selected at random from among the first K units in the frame, and every Kth unit thereafter is selected into the sample. We note, therefore, that there are at most K different samples possible using this scheme, and in the simplest situation, each of these K samples has the same chance of being selected. The random nature of this type of sample is achieved through the selection of the first unit at random - for example, by the use of a random number table or other such random device. A sample of this kind is referred to as a <u>1 in K systematic random sample</u>, and K is called the <u>sampling interval</u>.

There are two situations to be considered in selecting a sample of a given size, say n, from the population of size N.

Let us suppose that the population size is an exact multiple of the desired sample size, i.e., N = k.n. The population may thus be divided into n groups each of size K, and by selecting a random starting number between 1 and K and choosing the unit corresponding to the selected number and every Kth unit thereafter, our sample size n is determined.

If, however, the population size is not an exact multiple of the desired sample size, we often choose as the sampling interval K, the whole number either immediately above or below the ratio $\frac{N}{n}$. The resulting sample will not be exactly of size n but will be either slightly larger, or smaller.

An alternative to this device is to consider the population to be arranged in a circle. The starting position is determined, at random, by selecting a random number between 1 and N and selecting the corresponding unit, and then every Kth unit thereafter in a circular fashion until n units have been selected.

The sample consists of units U_9, U_{13}, U_2 , U_6 .

A prime advantage of this system over simple random sampling is the ease with which the sample can be selected. Often the selection procedure takes substantially less time and results in fewer mistakes in the identification of selected units. Additionally, the systematic sample tends to be spread more uniformly over the population and unusual samples with a preponderance of units from one region, for instance, are avoided.

A major disadvantage to the system stems from the systematic nature of the sampling scheme. Should there be a cyclical, or periodic, property to the ordered units, and the sampling interval, K, corresponds to the length of the periodicity, the resulting sample will provide very poor results. Indeed in the extreme case, the sample size n will provide no more information than a sample of size one.

For example, and this is a bit unrealistic, assume we wished to sample telephone calls emerging from a particular exchange on a weekly basis, but only selected one day of the week. Telephone traffic definitely has a weekly cycle, and if K=7 and we selected Sunday (at random) then a very different picture would emerge from this survey than had we selected Monday as the starting point. If K were some multiple of 7 say 14, 21 etc., the problem would persist. Often the periodicity in a population may not be known and using a form of systematic respect to their summer job earnings. Recognizing that the size of earnings may be influenced by the area of study, the population of students is subdivided according to the faculty of registration; Arts, Science, Education, Engineering, etc.

The selection of a sample of size n from the population is accomplished by selecting n_1 units from stratum 1, n_2 units from stratum 2, and so on, through N_L units from stratum L, such that $n_1 + n_2 + \ldots + n_L = n$. Within each stratum, the units are selected according to a probability sampling scheme for example SRSWOR, or systematic random sampling, or indeed any of the schemes still to be considered. While there may be some advantage to using the same scheme in each stratum, there is no necessity to do so.

This system of selecting random samples from individual strata is known as <u>Stratified Random Sampling</u>.

There are several advantages to the use of stratification in a sample survey.

1. It has been shown that when strata are formed in such a way that homogeneity of units within a stratum is achieved while heterogeneity between strata is also created, the overall precision of the estimates is vastly improved over that obtained without stratification with the same size of sample.

2. The use of geographical areas for stratification purposes may result in administrative savings in that fieldwork may be concentrated in areas resulting in less travel time for interview staff and for localized administrative control.

3. The nature of the units in a particular stratum may be such as to lend themselves to a sampling scheme quite different from that appropriate in other strata. This is easily accommodated.

3.6.4 Unequal Probability Sampling

In the general description of a probability sample it was specified that the units were to be selected using a scheme which assigned known probabilities of selection to each unit in the population. In the simple random sampling schemes, each unit in the population is assigned the same probability of selection. In addition, the systematic random sampling scheme has a similar property of equal selection probabilities per unit.

Let us consider a hypothetical situation in which it is desired to estimate the total production of fruit produced by fruit farms in a region. An examination of the population units, the fruit farms, reveals that the farms vary greatly in size with respect to the number of fruit trees on the farm: there are several very large farms and a few relatively small ones as well. In attempting to estimate the total production it may be that one would wish to include the larger farms in the sample, or at lest try to ensure their representation, as their production certainly will have a substantial influence on the total production in the region.

To accomplish this inclusion of the large farms, a stratified random sampling scheme could be used to advantage, with stratification carved out according to the size of the farms. Another potential procedure would be to assign the large farms a somewhat larger selection probability, and assign smaller selection probabilities to the relatively smaller units in the population.

A scheme which assigns different probabilities of selection to the units in the population is known as an Unequal Probability Sampling Scheme. In many cases, as in our example, the important aspect of the units is their size, with the larger units to receive higher probabilities of selection. This consideration of size has resulted in this type of sampling being referred to as Probability Proportional To Size Sampling or P.P.S. Sampling.

Number 2 is the Cumulative Total number of trees from 86 to 200 and because the first random number, 168, falls within the range of the associated numbers for Farm Number 2, this particular farm has been selected. In similar fashion 326 corresponds to a number associated with Farm Number 4, 771 with Farm Number 6, and 1315 with Farm Number 10. Our sample selected with P.P.S. sampling consists of farms 2, 4, 6, and 10. We note that the probability of selection of Farm Number 10, for instance, is the ratio of the number of random numbers (trees) associated with that farm (401) to the total number of eligible random numbers (trees) for the population (1905). The probability is thus 401/1905/ or 0.21. In simple random sampling the probability of Farm Number 10 being selected would be 1/12 or 0.083.

Clearly this process of selection would be exceptionally tedious if the population were large or the measurement of size large. The scheme requires the determination of the set of cumulative size measurements, a task which, though simple, is laborious. A somewhat more economical procedure is available however.

2. The second selection procedure to be described was developed by Labiri (1951). In this procedure, pairs of random numbers are selected, one associated with the unit number and the second random number with the measurement of size.

To illustrate, we consider again our hypothetical example. Because there are twelve farms and the maximum size of the farm is 401 trees, the first random number (r) in the pair is selected between 1 and 12, and the second random number (s) between 1 and 401. The number (r) identifies a farm which becomes therefore a potential member of the sample. If the

These two procedures are not the only ones available for the selection of a P.P.S. sample but are presented simply to illustrate the differences between this type of sampling and the equal probability procedure.

The P.P.S. sampling schemes do give increased efficiency to the estimation of the population parameters. The increase in efficiency will be particularly large if:

1. The population units vary substantially in size;
2. There is a measurement of size which is highly related to the variable of interest in the survey;
3. The measurement of size is available and reliable.

The use of P.P.S. sampling brings with its use, however, added complications:

1. The selection procedures tend to be complicated;
2. The estimation procedures are more difficult as they must reflect the unequal probabilities with which units are selected;
3. Generally speaking, the use of Without Replacement P.P.S. sampling causes very special problems both in selection and in estimation. As a result of these problems, very often with replacement sampling allowing for repetition of units is done.

As a final comment, it should be stressed that further refinements to the sampling procedure can be achieved through the combination of stratification of the population followed by P.P.S. sampling within the defined strata. It should now be possible to see the beginnings of the complexity that is possible in many sampling designs.

Secondly, because the units upon which measurements are to be made are clustered together, it may prove to be less costly to obtain the mesurements by virtue of such considerations as reduced travel time and decreased administrative complications.

If one were free to form clusters of units at will, the principle upon which they would be constructed, so as to give a high degree of efficiency to our estimation procedure, would be to form clusters of dissimilar or heterogeneous units. The selection then of a cluster would result in a wide range of units being measured giving broad coverage. Clusters of highly similar units provide less information about the measurement of interest. Usually, of course, it is not possible to have such freedom to artificially create clusters and, if attempting to decide on the use of a cluster sampling procedure, one would look at the naturally suggested clusters to see if they appear to be internally heterogeneous. In this case, cluster sampling usually provides excellent results. Unfortunately, it would seem highly likely that there will be a tendency for the clusters occurring naturally to consist of units with substantial similarity. In such cases it is suggested that one choose a large number of small size clusters rather than a small number of large clusters.

At first glance, it may appear that cluster sampling and stratified sampling are essentially the same. On closer examination, it is seen that this is fundamentally not true. There are striking differences:

1. In stratified sampling, samples of units are selected from each and every stratum whereas in cluster sampling only a sample of clusters is selected.
2. In stratified sampling, the units are arranged so that a stratum contains homogeneous units. In contrast, cluster sampling efficiency is increased if the units in the cluster are not homogeneous.

The estimation procedure, naturally, will not be simple given the complexity of a multi-stage sampling scheme. However, with the knowledge of the sampling procedure, and the selection probabilities, the estimates of the population parameter and their estimated standard errors can be built up stage-by-stage in a well-defined fashion.

Together with the expected decrease in cost from such sampling, a primary advantage of multi-stage sampling is the likely ease with which a frame for the various stages can be obtained. In the multi-stage example, for instance, the ultimate frame of households consists only of those located on the selected city blocks and there is, thus, no need to obtain a listing of all households in the entire country.

3.6.7 Some Additional Refinements

3. Multi-Phase Sampling

In many situations rather detailed information is required on the units in the population. Some of this information may be expensive to obtain, or present some peculiar difficulties. It is often advantageous to select a large sample of units from the population and obtain the bulk of the information required.

For those respondents which require special techniques or considerations, or are costly to interview, a second, likely small sample is selected from among those units previously selected from the population. The units selected in the sub-sample are then more extensively examined for their characteristics.

If the process involves a first sample followed by a second sub-sample, the procedure is known as two-phase or double sampling. Of course, the idea can be extended to many phases.

To illustrate, if the item of interest were the total production, Y, of wheat in a region in a particular year, the knowledge of the production of wheat for a sample of farms in the particular year, y, and in the previous year, x, together with the total production, X, in the region for the previous year can be used to provide an improved estimate of the total production in the current year, $\hat{Y} = \frac{y}{x} \cdot X$.

6. Domain Estimation

There is often interest not only in producing parameter estimates for the entire population, but also for well-defined sub-populations within the population. When this is the case, the sub-populations are referred to as domains, and are, themselves, treated as populations for study.

It must be stressed that if estimates are required for domains a sufficiently large sample must be selected from the domain in order to achieve the desired results. This may result in the overall sample size required for the entire population being very large.

7. Sample Size

In order to determine the appropriate size of sample for any given sample survey, a substantial amount of information about the population is required. In particular, it is necessary, in most cases, to have information about the variability inherent in the population - information which so often is simply not available.

In addition, there must be a specification made of the degree of precision which is desired in the estimator. In other words, how large an error of estimate is permissible. But this, still, is not sufficient, as there can be no guarantee that a given sample will produce an estimate within the required bound. Together with this error specification, Examination of the table of sample sizes reveals that the demand for greater confidence in our sample estimate results in the requirement for increased sample size. Also, if we require that the error of estimate be smaller, the sample size must be increased. The most striking result, however, is the relatively small influence that the population size has on the sample size requirements. An increase, in the population from size 1,000 to size 500,000, a 500 fold increase, results in the need to increase the sample by a factor of approximately 1.5 at most.

It should also be noted that the population variability is extremely important in the determination of the sample size. If we now consider a doubling of the standard deviation of the population from 25 to 50, the resulting sample sizes, with d = 2 at 95% confidence, are given below.

N	n
1,000	706
2,500	1225
5,000	1623
10,000	1937
50,000	2291
100,000	2345
500,000	2390
2,000,000	2399

It is clear that for the large population sizes the increase of the population variability by a factor of two results in approximately a four-fold increase in the required sample size.

The use of multiple mailings, call-backs, and subsequent in-person interviews is common in achieving high response rates.

Should the survey involve the asking of sensitive questions of the selected units, it is expected that the non-response rate will be increased. Chapter 6 below contains additional procedures for coping with non-response in the analysis of the data.

3.7 CONCLUSION

The subject of sampling is vast and complex. This chapter has surveyed only the most commonly applied procedures. The next chapter addresses the vital area of survey management. Although a sigh of relief may be permissible once the questionnaire and sample have been selected, the effective supervision of field work, data management, and quality control are essential to controlling the costs of the survey project.

Chapter 4

MANAGING SURVEY RESEARCH

4.1 INTRODUCTION

All too frequently, the management of a survey is secondary to issues such as sampling and questionnaire design. Yet poor management of field operations, data entry, and data documentation can easily negate the other aspects of survey design, no matter how well executed.

This chapter reviews the major issues in survey management, such as selecting and training interviewers, implementing the questionnaire (field operations), and data management. The last topic is particularly important, for surveys often can be used for analysis of other than the immediate question. A well designed and documented data file may enable an agency to reuse the information at some future point, thereby avoiding the costs of another survey. Also, the analysis of a survey may be interrupted for a variety of reasons, and a well documented file greatly eases the "re-entry" costs when the project is resumed.

4.2 SELECTING AND TRAINING INTERVIEWERS

Much folklore attends the selection of interviewers. Perhaps the most common is that middle-aged women make the best interviewers and field managers. Other often repeated dicta include not hiring young students, and matching the race, gender, and age of the interviewer with the respondent. In general, however, it is possible to identify certain attributes of an interviewer which are important, and which are not, in principle, related to gender, race or age. Of

The second phase is the actual job interview. Aside from the usual criteria required by all employers who seek employees who will have a high public exposure (clarity of thought and expression, and general deportment), it is important to examine the candidate for specific abilities to ask questions. Thus, the interview which seeks to evaluate a candidate for his/her potential as an interviewer must be structured so as to reveal these capabilities. A useful tactic is to explain little about the tasks, but to have the candidate probe for employment particulars. As with job interviews in general, it is useful to let the candidate do most of the talking.

The matter of deportment is not solely of concern to a finishing school, but indicative of the receptiveness with which the respondent may give the interviewer. A professional appearance that is not intimidating is very important. Respondents are more likely to feel secure that the information they provide will be responsibly handled if the interviewer provides the impression of a competent professional. For a telephone interview, dress is not important, but a telephone "presence" is. A useful procedure for screening applicants for telephone interviewers is to conduct the job interview on the phone.

Once the "short" list of interviewers has been prepared, the final selection phase for any given survey merges with the training and pilot study stages of the overall project. Frequently, interviewers are more suitable in some social settings than others, and during the training and pilot study stage may self-select themselves out of the project. For example, an interviewer who has no difficulty in working on a survey of home heating may be uncomfortable in dealing with personal questions. At the pilot study stage, the project director must be especially alert to any potential deficiencies which may disqualify an interviewer from any particular survey, for carelessness here can lead to serious distortion for a significant portion of the sample.

perience of many other researchers. The only caveat is that all scientists must cultivate the capacity to be surprised - men can make good interviewers!

4.2.2 Interviewer Training

It is often convenient, especially given budget pressures, to assume that interviewing requires no training period. This is false. Every questionnaire, no matter how straightforward it appears, has subtleties; if it does not, one can legitimately ask why it is being done in the first place. Even simple opinion polling requires care.

To reduce costs, it is often recommended that the interview training process and pilot study phases be combined. In the first set of interviews, the project director and field supervisor should accompany the interviewer (or listen in on a telephone survey), and subsequently provide a critique of performance. Those who perform well from the start can be allowed to proceed unaccompanied to other interviews in the pilot study, while others may require further support. It is at this stage that inappropriate interviewers must be dropped from the project or reassigned to other duties.

For agencies which contemplate an ongoing need for survey research for a major project, there is merit in a more formal training function. Here, actual classroom instruction, supplemented by video taping and group criticism of performance, is very valuable. If properly handled, such group criticism can lead to improved morale among interviewers and instill in them a greater sense of the worth of the project. This has important impacts upon the field operations stage discussed below.

Second, especially when surveying a heterogeneous group, matching the interviewer and respondent can be expensive or impossible.

Third, even if in principle one could precisely match the interviewers and respondents by a number of criteria (age, race, gender, income, education, etc.) why would one interview the sample respondents? Why not just survey the interviewers?

In summary, the whole question of matching interviewer characteristics and respondent characteristics must be approached with extreme caution. In those situations where it appears as though race and gender (or any other immediately identifiable attribute) of the interviewer may cause an adverse reaction on the part of the respondent, the objectives of the survey should be carefully reconsidered. Only after the pilot study has clearly identified an adverse respondent reaction, based upon interviewer characteristics, should any attempt be made to match the two parties. Even here, the perceived need to match often reflects poor interviewer hiring. In certain settings such as interviewing core area residents or prison inmates, women may be at risk. Here teams composed of male and female interviewers should be considered.

4.3 IMPLEMENTING THE MAILOUT SURVEY

Many survey researchers endure response rates of 10 to 15 percent with fatalism. They feel that because a third party (the postal system) is involved and because the procedure is cheap (or so it is imagined), little can be done to enhance response. In other words, mailouts are useful for "quick and dirty" analyses usually to be undertaken when the client is somewhat uninformed about the mechanics of quality control in survey research. This thinking is quite wrong. Through careful planning, mailout questionnaires can enjoy very respectable revidual responses are valuable and indicate how the names and addresses were obtained. A sample covering letter is shown in Figure 1. To reiterate, the covering letter requires careful thought and planning, and should not be an afterthought at the last moment.

Figure 1

THE UNIVERSITY OF MANITOBA

THE INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH
Faculty of Arts

Winnipeg, Manitoba
Canada R3T 2N2
(204) 474-9422

Dear Householder:

The Institute for Social and Economic Research is preparing a study for the Department of Energy, Mines and Resources on energy conservation and fuel use by homeowners. Your name was one of only five hundred drawn at random by a computer from the City of Winnipeg tax records for participation in this survey. Because we are asking only a few selected households to complete this questionnaire, your participation is very important.

Please complete the enclosed questionnaire and return it in the enclosed envelope. Once the computer file on this information has been completed, all reference to your name and address will be deleted and this questionnaire will be destroyed. Your name, address or any information you supply will not be provided to any individual,private companies or government agencies. All information gathered by surveys undertaken by the University of Manitoba is confidential and anonymous.

At the end of the summer you will receive a report which summarizes the information gathered in this survey. This report will also be sent to government officials and politicians at the provincial and federal level. By participating in this survey, you have a chance to express your opinion on energy policy to those currently involved in designing those programs.

If you have any questions about this survey or the questionnaire, please call 269-5536 during normal business hours, or between 6:00 p.m. and 9:00 p.m. Monday to Thursday.

Sincerely,

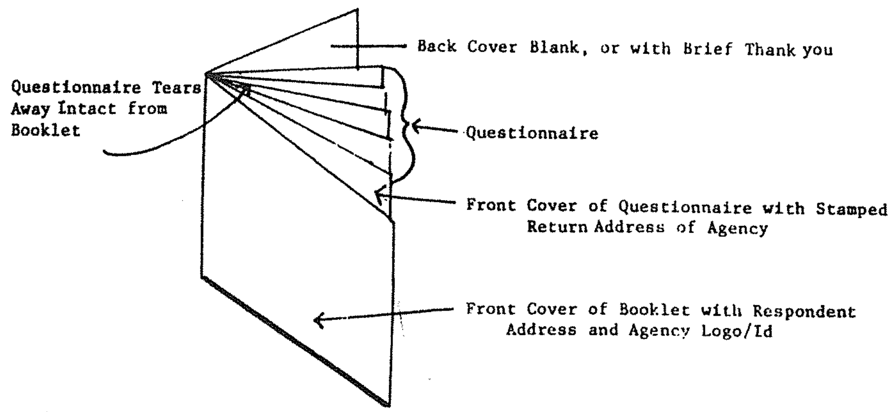
Greg Mason

Dr. Greg Mason
Director

GM/lt

enclosure

Figure 2

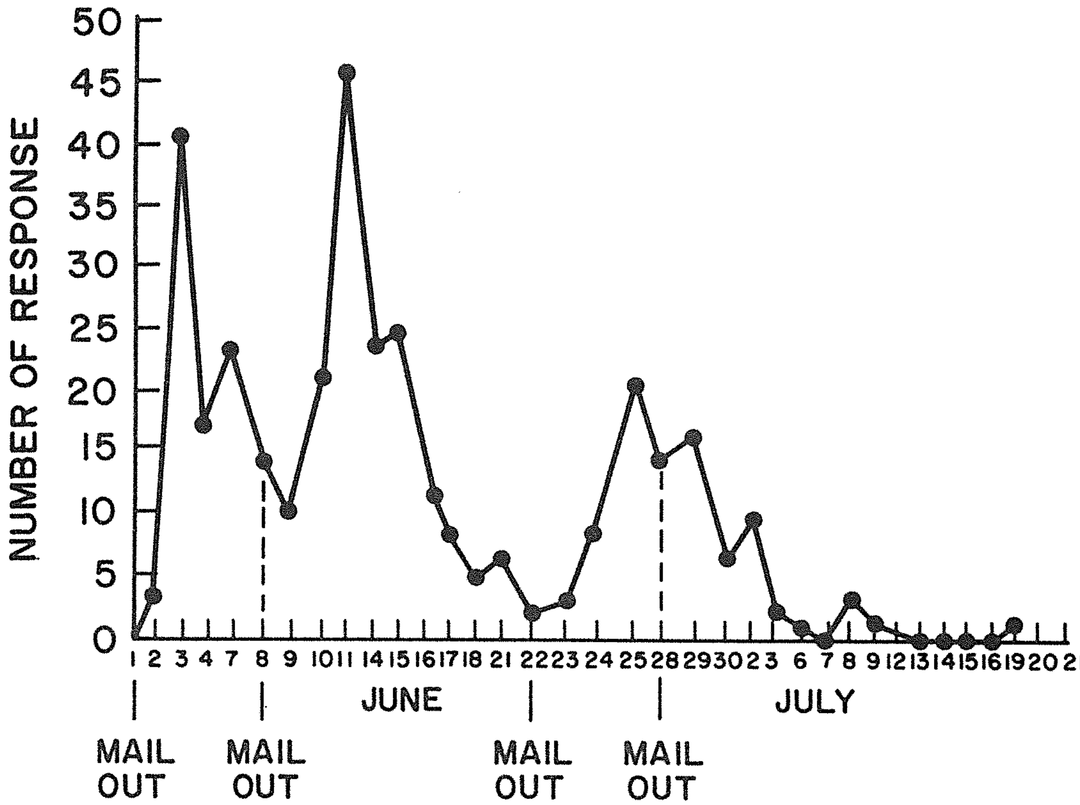


The return envelope should be numbered, (as should the questionnaire) thereby allowing the elimination of respondents from the list of those sampled. In some cases, a survey may seek to merge so-called "self report data" with administrative files. In this case, as the computer file is constructed a unique number must be assigned to unite the various files. It may be a realty tax number, or even the name of the respondent. This is permissible for the purpose of file construction, but all identifications should be dropped from the file once it has been completed. A certain proportion of respondents try to obliterate these identification numbers, and so stamping the inside of the return envelope is a useful dodge.

It is also wise to enclose a brief description of the agency. This provides the respondent with a more complete view of the organization, enhancing the importance and legitimacy of the survey.

it is assumed that little more can be done to enhance the response, but this is quite wrong. Figure 3 shows the typical response rate. Clearly, follow-up accounted for over 60 percent of the ultimate response.

Figure 3



SURVEY RESPONSE OVER TIME

Figure 5

THE UNIVERSITY OF MANITOBA

THE INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH
Faculty of Arts

Winnipeg, Manitoba
Canada R3T 2N2

261-5536

June 22, 1982

Dear Homeowner:

About three weeks ago I wrote to you seeking your opinion on energy policies for homeowners. As of today we have not yet received your completed questionnaire.

The Institute for Social and Economic Research has undertaken this study because of the belief that the opinions and lifestyles of Winnipeg homeowners should be taken into account in the formation of public policies affecting energy consumption and conservation.

I am writing to you again beacause of the significance each questionnaire has to the usefulness of this study. Your name was chosen through a scientific sampling process in which every homeowner in Winnipeg had an equal chance of being selected. In order for the results of the survey to be truly representative of the opinions and actions of Winnipeg homeowners, it is essential that each person in the sample return their questionnaire.

In the event that your questionnaire has been misplaced, a replacement is enclosed.

Your co-operation is greatly appreciated.

Sincerely,

Donna Brown

Donna Brown
Project Director

DB/lt

enclosure

tive respondent called in to indicate that they were not going to reply because a television show the night before had alleged that all data in computer files could be easily accessed by anyone. After a ten minute coversation in which it was explained to them that even if someone actually looked at the file, no identification of the respondent was possible, they did indicate they would reply, although this could never actually be checked. Fear of modern computing procedures are real (as are the dangers from misusing this technology) and care must be taken to reassure respondents in this regard.

4.3.6 Encouraging Response

Some survey researchers use a financial reward such as enclosing a small amount of money with the questionnaire, or indicating that lottery tickets will be sent upon completion of the questionnaire. Although there is merit in compensation, care must be taken not to antagonize respondents. Few research organizations can afford to pay anything but a token amount. Respondents who are motivated by pecuniary incentives could merely complete the questionnaire as rapidly as possible to qualify. In general, a survey with clearly defined objectives, which provides insight to the respondent (in a summary report) and appeals to altruistic motives produces more accurate data than surveys which rely upon token payments.

There is considerable controversy in this area of survey logistics, and proponents of each view tend to be convinced of the correctness of their position. If one can afford a significant incentive (e.g., a major appliance) then payment for completion may have potential. However, enclosing a dollar to compensate for the time spent in completing a survey, is not clearly superior to providing a summary report to satisfy natural curiosity.

A drawback is that telphone listings may not cover the population discussed, although this is much less so now than twenty years ago, and much less so in North America than Europe. Certain groups are almost always missed such as Hutterites and native Indians on reservations. It is quite difficult to verify the coverage of the population from telephone listings and telephone samples tend to be the weakest when analyzing non-response.

Rather than counting actual listings between sample points, it is possible to use a ruler to approximate the spacing. For directories that have a fixed number of entries per centimetre, this can be quite acceptable. Even if there is more variation, the error is usually slight and will generally be random.

When a particular listing is not a residence, one can use a simple rule, such as alternating between the next residential listing above and below the non-residential number selected. In some cases, where a listing occupies two lines (such as professionals), they are selected only if the count (or ruler) falls on the residential number, otherwise they are treated as a non-residential number. There is a small bias towards selecting households which may be important in some surveys.

The second procedure is <u>random digit dialing</u>. One variant of this procedure calls for the identification of all three digit numbers representing telephone exchanges in the area of interest. Next, a table of random numbers or a random number computer program is used to generate the remaining four digit number. For areas with a large number of exchanges, a random selection of these can be undertaken as a first step. Finally, the interviewing proceeds after discarding those that are not working, or non-residential as provided by the phone company.

The single major drawback to this approach is the difficulty of eliminating non-residential and unused numbers. Often a major subset of an exchange may ei-

4.4.2 <u>The Information Letter</u>

Given the increasing use of telephones for promotion, it is unwise to contact the respondent without warning, except on short opinion polls. Even when the target group is well identified, such as professionals, members of an organization, students, etc., and response rates generally high, the use of an information letter is good practice. It resembles the covering letter used in the mailout, but must indicate generally when the telephone call will be made and the appropriate time required. If costs and time permit, it is wise to indicate that a call will be made to arrange an appointment. This is especially useful if the questionnaire is long (over 15 minutes) or one which requires some background research on the part of the respondent.

4.4.3 <u>The Introduction</u>

The initial impression made by the interviewer on the phone is crucial. The caller should identify themselves, recall the letter sent out and <u>briefly</u> review the purposes of the study. Stress should be placed on the importance of the study and that all information will be confidential and anonymous. The caller should then inquire whether it is convenient to conduct the interview; if not, an appointment should be made to call back later. It is also very helpful to indicate how long the survey will take.

4.4.4 <u>Conducting the Interview</u>

In many cases, respondents will wish to know more about the survey. Interviewers should be prepared to answer clearly queries about the agency conducting the survey, for whom the survey is being conducted, what are the purposes of the survey, who is the principal investigator, how was the respondent's name ob-

4.4.5 Monitoring the Interview

It is most desirable to randomly monitor telephone interviews for two reasons. First, valuable assistance can be rendered to the interviewers so that performance can be improved. Second, it reduces the incidence of faked interviews. For this reason telephone interviews should always be conducted out of a central facility, except in rare circumstances (i.e., extreme deadline pressure) should interviewers be allowed to conduct a survey from their own homes or an unsupervised office. Monitoring can be accomplished quite simply by using duplex plug-in jacks and then inserting the second line into a receiver. Removing the mouthpiece eliminates any noise that may be inadvertently caused by the supervisor. Most professional survey research firms have quite elaborate facilities to accomplish telephone interviews, where interviews can even be recorded for subsequent verification of responses.

4.5 IMPLEMENTING THE IN-PERSON SURVEY

Many of the general comments about mail and telephone surveys pertain to face-to-face surveys. The initial letter, presentation of the project and the interaction between interviewer and respondent are all relevant. One aspect, monitoring interviewers performance, does deserve some additional comment.

It is the responsibility of the supervisor to monitor and appraise the performance of the interviewers. Very early in the survey, some norms of performance will be established. It is important to spread the difficult interviews among the field staff so as not to unduly burden any portion of the staff (yet another argument for not matching interviewer and respondent). When certain interviewers are failing to meet these norms, the supervisor must immediately determine why, and rectify the situation even if it means termination. The integconditions of work must be reviewed to ensure this possibility exists. Instances of data falsification and other violations must be clearly documented and the manager should review the grounds for dismissal with the relevant labour relations officers. This especially is important in the public services where survey employees may be covered by quite strict labour agreements.

In other situations survey researchers may be part-time or casual employees (e.g., summer students). In these cases a contract is essential; an example is provided in an appendix. Of course these general remarks on interviewer and employee relations are pertinent for other aspects of survey research. It is much easier to cope with potential problems in advance rather than attempting to deal with an employee relations problem in the midst of field operations.

4.6 GENERAL LOGISTICAL CONSIDERATIONS

To summarize, it is easy to dismiss the administration of survey research. This is a particular danger for agencies which elect to do a survey "in-house." The low response rates and low quality of results that seems to bedevil much survey research, is frequently attributable to weak management. It is wise not to allow senior management to authorize an in-house survey, unless they are prepared to allow full-time supervision by the staff. The supposed economies from doing survey research in-house quickly evaporate when shoddy results are realized. Agencies which specialize in survey research, often have several distinct advantages in the execution of survey research, and despite their apparent higher cost, they should not be rejected until it is clearly demonstrated that sufficient internal resources will be made available to execute the project.

All surveys must run out of a central office. This may be merely a desk, phone and secure filing cabinet, or it may be more elaborate. For telephone

Figure 6

Sample Logging Sheet

CONTACT DIARY
Anywhere, Social Services Survey
1982

Respondent Number ___ - ______

Name ______________________

Address ______________________

____

Telephone ______________________

Contact Diary

	Date	Result
First Contact	______	____
Second Contact	______	____
Third Contact	______	____

Interview Date(s) ______ ______ ______

Interview Completion Date ____________

Reasons for Refusal ____

Respondent Data House Value ______

Age of Respondent _____

Number of Bedrooms _____

Occupation ______

Codes
- NA - No Answer
- R - Refusal
- A - Agreement to Participate
- TB - Too Busy
- IP - Information Required Too Personal
- MT - Mistrust of Procedures, Computerized Data
- TO - Too Old, Sick or Disabled
- LP - Language Problems
- MV - Moved

Figure 7

Q.C.

15. What is the number of the local union you belong to?

JOB 1__________

JOB 2__________

JOB 3__________

J1 ☐ J2 ☐ J3 ☐

16. Do you still have this job with (NAME OF EMPLOYER)?

1 YES → GO TO Q. 24

2 NO

J1 ☐ J2 ☐ J3 ☐

17. Please tell me the date when you left this job.

RECORD DAY, MONTH, YEAR

J1 __/__/__ DD MM YY J2 __/__/__ DD MM YY J3 __/__/__ DD MM YY

18. Have you received any severance pay since (DATE OF LAST INTERVIEW)?

1 YES

2 NO → GO TO Q. 20

J1 ☐ J2 ☐ J3 ☐

19. How much severance pay have you recived since (DATE OF LAST INTERVIEW)?

J1 $______ J2 $______ J3 $______

4.7.2 Creating the Computer File

Many survey researchers are marvellously skilled at research design, superb field managers and astute statisticians, but not capable when it comes to the mundane task of transferring data to an electronic file. There are several simple precautions which can considerably alleviate problems at this stage.

The first point to be made is that data entry at a video display terminal (VDT) is boring, routine and dull. It is useful to combine the tasks of interviewing, questionnaire checking and data entry. This allows the staff to obtain a better perspective on the work, and leads to improved productivity.

Second, it is most common to use VDT's (rather than cards) to enter data. Some attention to layout is important. For example, it is very worthwhile to create a coding sequence which leaves spaces between the variables generated as shown in Figure 8 above. This may seem a minor point, but it does assist in visual inspection of the records on the VDT screen as shown below in Figure 9 where two record lines, one "open" the other "closed", are displayed. Clearly, the open format is

Figure 9

```
12 16000 1 130 46 17 23      (open record)

1216000130461723            (closed record)
```

much easier to comprehend and each variable stands out.

The typical screen is set up to permit either 80 or 132 characters. Assuming an average of three characters per variable, plus a space, there is room for at most 20 or 33 variables per record line, less than generated by the typical survey. In the era of cards, the fact that only 20 variables per record (card)

Figure 11

File A:

001	13	14.65	1	0	1	6500	25	......	65
002	13	12.72	0	1	0	7500	14	......	61
113	16	9.71	1	1	1	8520	9	......	72
......									
......									

File B:

001	2	1360	1	0	1	1	0	1	1	0	31	60	...13
002	1	1100	1	0	1	0	0	1	1	0	61	35	...14
......													
......													

File C:

001	3	1	1	0	0	1	0	1	3	4	1	7	13
002	1300	26573	16	35	0	1	1	1	1	7 ...	8		
......													
......													

Note that card numbers are no longer required. Each file could have a theme reflecting different parts of the questionnaire, and it is not necessary to fill the entire record line. Each variable is identified clearly within the record,

1. <u>Consistency</u>. An example clarifies this. It is clearly illogical to record a man as having given birth to children. On a long record, such a contingency might not be easily spotted, especially if there are many variables. Neither the response (male) nor the number of children are unacceptable values individually. In combination however, the data makes no sense.

2. <u>Range</u>. Some variable should <u>probably</u> not lie within certain ranges. For example it is probably very rare to find anyone with an income in excess of $50,000 in a survey of the urban core. Although not impossible, the low probability of such an occurrence should serve to signal a check of the original survey forms. This form of checking is not as precise as validity checking since there is clearly room for discretion in setting a cut-off value for the variable in question.

3. <u>Range</u> <u>and</u> <u>Consistency</u>. The final form of validation combines both a range and consistency check. For example, it would be rare to find an unemployed person with high income. Again the decision as to the cut-off level for income is judgmental.

These validation checks are best implemented in the context of the computer program used to undertake the analysis. Most modern statistical software contains logical "if" statements that permit the printing of records with wild codes, out of range values, and inconsistencies. These checks are most valuable in longitudinal studies where time dependent variables may be inconsistent between panels. Cross-sectional data may pass all validation checks, yet when several surveys are compared, inconsistencies may become quite obvious. For example, in a longitudinal panel study of job holding, one variable may indicate no change of job over several surveys, yet occupational codes may change. With-

4.8.1 Record Format

The record format is little more than a specification of the layout of the machine readable records. This can be very simple and consist of a simple listing of variables along with specification of the columns in which they occur as shown below:

TABLE 1

Variable	Column
Response	1 - 6
Age	8 - 9
Income	11 - 17
...	...
...	...

Although such a simple specification may suffice for the primary research team it has little merit in terms of making the information available for secondary analysis. Even the survey designers may imperfectly recall what each variable actually measures if they only have the mnemonic to go on after a year of absense from active use of the information.

4.8.2 Codebook

More useful is a codebook, comprised of the questionnaire, specification of how each variable is generated in the questionnaire, and notes on the special interpretation of variables especially of variables which have been recoded or assembled from qualitative information obtained by the interviewer.

Telephone Surveys

Do	Don't
1. Centralize the telephone office.	1. Call during the supper hour (5:30 p.m. - 7:00 p.m.).
2. Change the questionnaire so that it reads easily on the phone.	2. Call before 9:00 a.m. or after 10:00 p.m. unless by appointment.
3. Train the interviewers using "dry" runs.	3. Press a hostile respondent.
4. Use an advance letter to introduce the survey and its purpose.	4. Allow the survey to be conducted from the interviewer's home phone.
5. Provide the interviewer with an operations manual.	5. Let the questionnaire leave the office.
6. Provide the interviewer with a set of expected questions and suggested answers.	6. Tolerate any breach of confidentiality.
7. Train the interviewers to be sensitive to the form of the response as well as its content.	
8. Use contracts for interviewers which clearly explain the responsibilities of all parties.	

Chapter 5

BUDGETING SURVEY RESEARCH

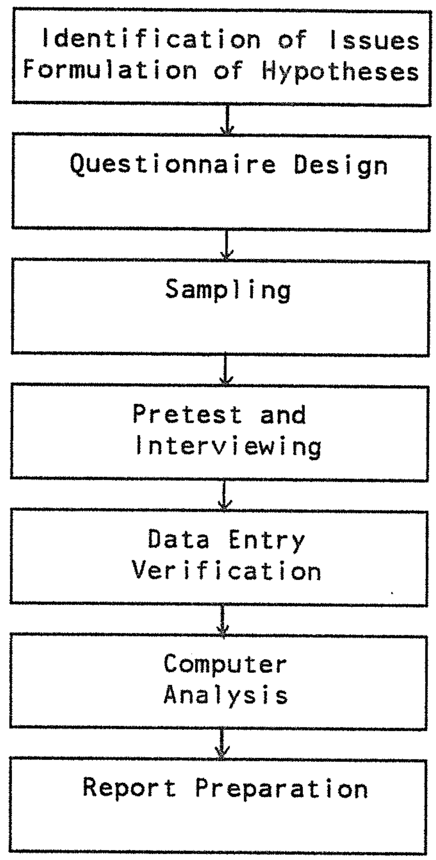
5.1 INTRODUCTION

Survey research is expensive. Many investigators grossly underestimate the costs of acquiring information in this manner, and even modest surveys can produce serious cost overruns unless care and realism are applied in liberal doses. In the face of budget constraints handed down by senior managers, survey researchers can be tempted to compromise. Most surveys are reasonably robust with respect to sample size (although care must be taken not to preclude analysis of special subgroups within the main sample frame by limiting sample size too severely) however, other aspects of a survey may not be so resiliant. For example, mailout surveys quickly disintegrate when follow-ups are removed. Similarly, telephone interviews become difficult to complete if there is no covering letter provided for respondents who are expected to be able to recall specific details such as profits for a manufacturing firm over the past ten years.

This chapter examines the major budgetary considerations for the three types of surveys, identifies some false economies and considers how to impose some sensible cost control, especially on those aspects of a survey which may be undertaken "in-house". Also examined are the problems of budgeting time and the creation of reasonable deadlines for completion of the entire project. Surprisingly with some ingenuity it is possible to complete survey research in a short period without necessarily increasing costs beyond acceptable limits. Indeed the careful scheduling of the various steps in a survey can save considerable money.

Figure 1

Steps in Survey Research
(Naive Version)



5.2.1 Questionnaire Design

Questionnaire design will amost always take several weeks. In the development of the research hypotheses, or in the background research to the study, a literature review will undoubtedly have been done. Central to this activity is the identification of other surveys, germane to the field. The development of a questionnaire is greatly assisted if other instruments are obtained and reto extract certain specified numbers of households according to structure (numbers, marital status, etc.). Once this basic sample had been selected, the final stage involved a complex mathematical procedure (integer programming) which assigned various households to certain combinations of negative income tax policies (combinations of basic annual guarantee and level at which "extra" income would be taxed). The objective was to produce an experimental design that would yield maximum information for a given cost. The entire procedure took several months.

Frequently surveys can exploit an existing list, such as property tax records, motor vehicle registrations etc. If these are computer based, then sampling becomes quite simple. It is very useful to also simultaneously extract other information on the unit of analysis, as this is valuable in analysing response patterns and assessing the overall representativeness of the sample. In general, for most surveys, sampling will not occupy significant resources, especially if an existing computer based list can be utilized (and computing costs can be arranged by interdepartmental favour). Indeed, mailing labels can usually be developed from such lists, although some reprogramming is worthwhile to prepare labels which do not obviously look machine based.

5.2.3 Pretest and Interviewing

Pretest and interviewing consume about 50 percent of the total costs of a telephone and face-to-face survey. Even in a mailout, the questionnaire will usually be pretested using a face-to-face procedure and an *ad hoc* sample can be used, unless the sample quality is at issue and needs testing. Pretests are essential for all survey research. It is most disheartening to be stuck with a poor questionnaire or invalid sample and merely go through the paces since to that questionnaires will be completed other than with the respondent. For the typical interview, the current rate is anywhere between $12.00 and $17.00 per interview (1983). Assuming a completion rate of 4 interviews in an eight hour day this assumes a wage of $6.00 to $8.50 per hour. Depending upon the market conditions, these wage rates may or may not encourage qualified people to be interested in the job. Remember that most interviewers are part-time, with education above the norm, implying that they may have some discretion in whether to accept work. Payment of a salary or regular hourly wage encourages professionalism, but does place greater burden upon the field operations manager to ensure that productivity remains high.

Postage costs obviously are important for mailouts. Prepare the package and obtain a precise quote from the postal service for the cost; it will be considerably higher than the cost of mailing a single first class letter. If survey research is a regular feature of the organization's work, arrange for a special return stamp that can be printed on all material which will be returned. In this way, only those questionnaires which are actually returned will cost. For small mailouts (less than 200-400), pre-stamping all return envelopes is not excessively expensive, for the non-respondents will only waste about $50.00 or so. For larger mailouts, the costs of non-response can rise dramatically.

5.2.4 Data Entry

Data entry and verification are directly related to the number of questionnaires and their complexity. Interviews with a large fraction of open questions consume very large amounts of skilled staff time as these responses are recoded into digital formats. A pretest is invaluable in reducing the number of open responses. If the questionnaire must have such questions, it is legitimately survey researchers is that the computer charges they face as students bear little resemblance to the charges imposed by private sector firms. In some instances it may be possible for a public agency or crown corporation to slip the computing between other tasks undertaken by government computers, such as payroll or general accounting, and occasionally special deals can be established with "friendly" programmers to do this work. There are several potential problems to this course. First, most business oriented machines have only rudimentary capabilities to analyse social survey data. While there may be a version of a major "canned" program available, frequently it is an old version and one which may still contain bugs that have since been rectified in recent releases of that software. Second, most applications programmers do not really understand these programs. Not that these programs are especially difficult, but programmers typically are involved in writing quite specialized routines and have had little need to investigate statistical software. Third, the analysis of a survey will always take a distant last place to preparing the payroll (as it should) or undertaking other analysis. The computation of survey results for a remote research and planning division within government (and one which also may be responsible for program evaluation) may not be a priority for the computer division.

As an alternative, contracting the computer analysis with a private service firm does have the advantage that many of these organizations have sound expertise in statistical software. Typically they can prepare relevant analysis on time. The major caution is that the contractor must be thoroughly familiar with the procedures to avoid serious cost overrun. Especially when the service bureau is prepared to locate a remote terminal on site and instruct a staff member to access the system, there is a distinct possibility that an oversight will ment is that at least one person on the survey research team should be very familiar with the system in use. Also, it is essential that the computer programs and procedures be thoroughly documented to ensure that the work contracted was actually undertaken and to permit additional analysis in the future.

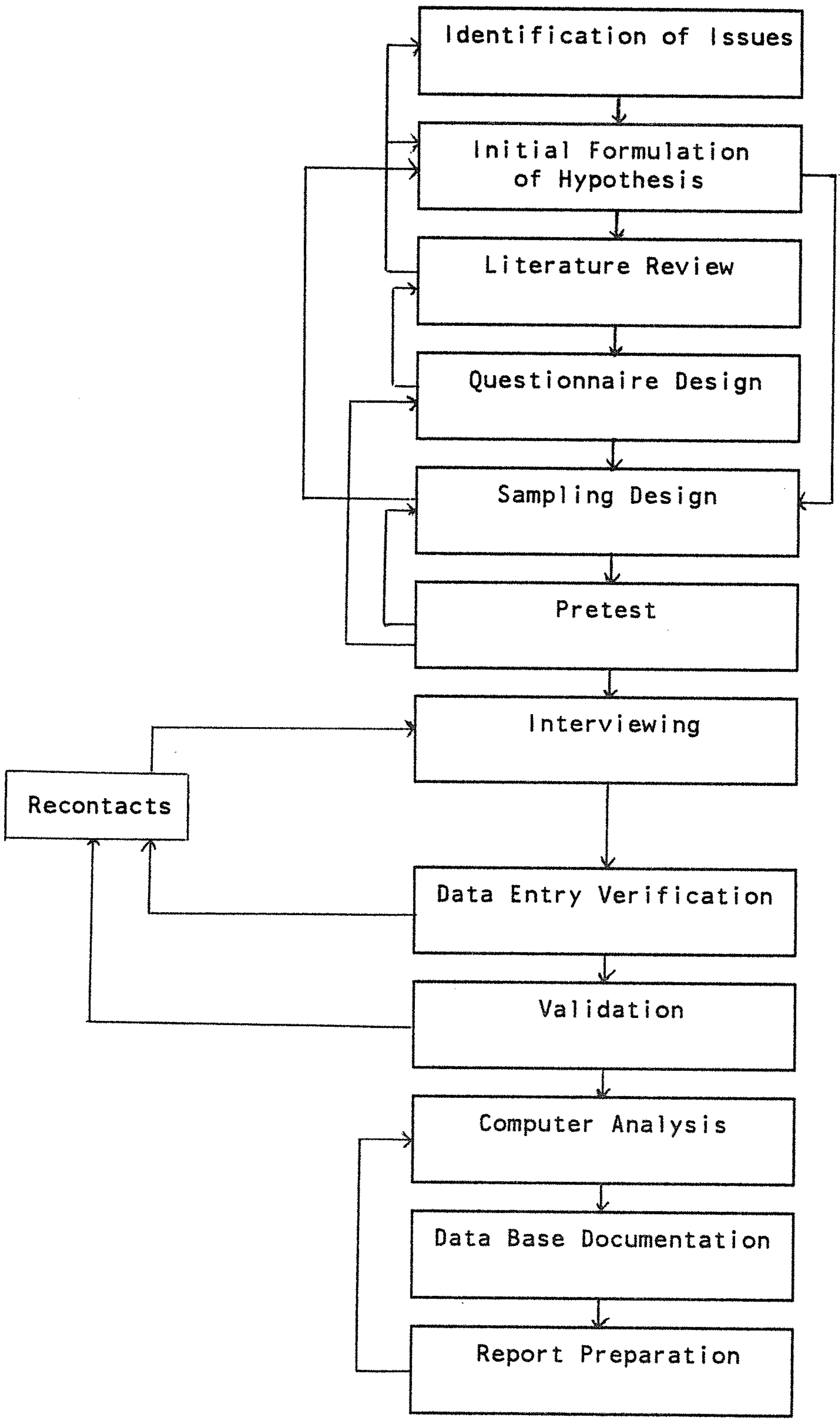
5.2.6 <u>The Final Report</u>

The final phase in the survey is the report. Most surveys will generate more than one report, but for the moment consider only the initial descriptive analysis. This consists of a general description of the objectives of the study, the methodology underlying the analysis, some explicit hypotheses, description of the sampling frame and field operations, and finally a presentation of simple descriptions of the data. These preliminary reports generally include frequency distributions for each question, as well as some selected cross-tabulations. The judicious use of graphs and charts is also strongly recommended.

These rather simple exercises serve a number of purposes. First, they become a very useful element in the overall documentation for the project. If a complete questionnaire is included (as it should be), as well as a code book, then the documentation is complete. A second purpose is that it provides a convenient form for presenting the main results to senior management. It is well to always remember the clients of research; providing a simple summary report can go a long way to allowing management to justify the expenditures, especially if the research yields interesting and useful insights. Finally, it is strongly recommended that the respondents receive at least an abridgement of the results. This produces much goodwill and assists those who must undertake surveys in the future. The promise of a summary report is often a strong inducement to participate if for no other reason than simple curiosity. A sample summary report is

Figure 2

What Actually Happens in Survey Research



data file need to be documented. It is not uncommon for large data bases to generate many subfiles. Each needs to be documented.

Finally, reports can (and usually will) produce unanswered questions. Hopefully this will not be due to poor initial design. The entire process can then recycle to the top as new objectives are identified, and new modes of analysis developed.

5.3 SOME SELECTED TOPICS

5.3.1 Allocating Time in Survey Research

A very useful planning tool for preparing a budget for a survey is the critical path as shown in Figure 3. This chart has several functions. First, it clearly places the survey in the context of the calendar year. If the survey intrudes upon other duties of the staff such as defending their department allocation in the annual budget sweepstakes, the critical path will demonstrate this and the survey should be abandoned, rescheduled, or let for external contract. A second feature of such a planning tool is that it clearly identifies the responsibilities of different personnel. The critical path also assists greatly in clarifying whether the survey is realistic; bottlenecks can be identified and adjusted. The survey team should all be instructed in the critical path, and each participant's role clearly explained. This is especially important for complex face-to-face interviews which ultimately depend upon the creation of a sound team.

Most survey projects last less than a year from the initial planning stage to the final report. It is also possible to completely execute a telephone survey on a narrow question (opinion on a proposed tax revision) within ten working days, although only experienced survey researchers ought to attempt this. A

Telephone surveys and face-to-face interviews can be varied by adjusting the number of interviewers. In general, ten is the maximum number of telephone interviewers which can be efficiently supervised and which can be housed in a reasonably large room (15 x 25) without incurring major renovation costs to install sound dampened partitions and significant telephone changes. Also, the management cost rises sharply as the face-to-face interview team rises beyond twenty. Finally, even given the part-time nature of interviewers work, it is unlikely that suitable interviewers will be attracted for periods of under two weeks.

5.3.2 Accounting for In-House Research

It is tempting to incorporate survey research within the overall activities of a research unit. By "hiding" the true costs of the research it may appear that the direct out-of-pocket costs are well within line. Indeed by allocating interview work among the secretarial staff, obtaining committments from computer centre personnel and slipping the questionnaire into the regular mail stream, the direct costs of the research can be very significantly reduced. There are two main arguments against this approach.

First, a research manager who must resort to this approach clearly demonstrates both that research has both relatively low priority within the division and to undertake such analysis is really very inexpensive. While it might be possible to execute such survey research one or two times at most, as a general rule it will prove impossible to undertake surveys on a continuous basis using in-house resources. When surveys begin to show progressively poorer returns, senior management will inquire why the quality of work is declining.

Second, imputing costs and actually increasing the burden of research staff may well compromise the objectives and performance of the unit as a whole. The months of May to mid-June. As it transpired, delays in the printing of the questionnaires pushed the survey back to June 1, which is the last acceptable date for releasing a mailout.

Budget of Mailout Survey as Prepared by The Institute for Social and Economic Research

1.	Field Supervisor 4 months @ $1,600.00/month		$6,400.00
2.	Data Entry 150 hours @ $7.50/hour		1,125.00
3.	Postage		
	a. First mailout 500 @ 45 cents/envelope	$ 225.00	
	b. Postcard follow-up 500 @ 30 cents/card	150.00	
	c. Second letter (special delivery) 250 @ $1.50/envelope	375.00	
	d. Final postcard 150 @ 30 cents/card	45.00	
	Total Postage		795.00
4.	Printing		
	a. Questionnaire 750 @ $1.00/form	750.00	
	b. Follow-up cards 650 @ 30 cents/card	195.00	
	c. Covering letters 750 @ 1.5 cents/letter	11.25	
	Total Printing		956.25
5.	Envelopes 1500 @ 5 cents/envelope		75.00

to prepare a list of victims and witnesses along with their telephone number. Due to the high mobility of the population, substantial recontacting efforts were required and the actual telephone interview period for a total sample of 250 was in the order of three months. Also, because many of the prospective respondents could not be located, it was deemed necessary to establish a minimum survey size. Finally, two separate questionnaires, one for victims, the other for witnesses were implemented.

Budget

Victim/Witness Survey
(December 1982)
The Institute for Social and Economic Research

1. Personnel		
a. Stipend for Principal Investigator 10 days @ $250/day	$ 2,500.00	
b. Senior Research Associate (Field Supervisor) 42 days @ $100/day	4,200.00	
c. Telephone Interviewers 3 @ 80 hrs. each @ $8/hour	1,920.00	
Total Personnel		$ 8,620.00
2. Temporary Telephones 3 @ $75 phone for 2 months		225.00
3. Computer Services 2 hours CPU time @ $300/hour		600.00
4. Data Entry 150 hours @ $7/hour		1,050.00
5. Printing 300 questionnaires @ $1.00/questionnaire		300.00
TOTAL COST		$ 10,795.00

retical model, and although the other two case studies have this feature, the Winnipeg Area Study most clearly exemplifies this.

Budget
Winnipeg Area Study 1983
(Proposed)
The Institute for Social and Economic Research

1. Personnel		
a. Principal Investigator (Project and Questionnaire Design, 1/3 release time from academic duties)	$ 11,000.00	
b. Field Co-ordinator (for 5 months)	7,000.00	
c. Assistants to Field Co-ordinators (350 hours @ $9.00/hour)	3,150.00	
d. Interviewers (700 interviews @ $16.00 interview)	11,200.00	
Total Personnel		$ 32,350.00
2. Data Entry and Computer		
a. Data Entry		
Key to Disk Operators (2 for 200 hours @ $7.50/hour)	3,000.00	
File Verification (200 hours @ $7.50/hour)	1,500.00	
b. Computer Costs		
CPU Time (3 hours @ $500.00 hour)	1,500.00	
Space Rental	500.00	
Programmer (110 hours @ $11.00 hour)	1,210.00	
Total Data Entry and Computer		$ 7,710.00
3. Typing and Printing		
a. Clerical		
Questionnaire and Report Typing 150 hours @ $7.50/hour	1,125.00	
b. Printing and Postage		
Questionnaire Printing (800 copies @ $1.50/copy)	1,200.00	

5.4.2 Costing and Productivity

Be realistic about costs. Agencies which do "in-house" survey research must cost internal resources at their opportunity cost (cost of not doing normal duties). Allocations for adequate support facilities, such as centralized survey offices are essential. Contracts for work outside must be let to tender, and should generally be given to firms which have had a successful record in these endeavors. Try to arrange for options to extend statistical analyses at reasonable rates. Most reputable firms will provide follow-up consulting services and nominal rates to encourage further business. Also, on contract work, the "deliverable" must include the final report, all computer programs, questionnaires and, most importantly, a properly documented machine readable data file (usually in tape format). This can preserve the data base for further analyses, thereby reducing downstream policy analysis and development.

5.4.3 Computer Programming

The myth still persists that all computer programmers are eccentric genuises which perform miracles at 4:00 a.m., as long as they are allowed to wear bizarre clothing and babble incoherently. In fact, most programming is mundane, requiring little original insight. Most survey research is executed within established canned computer packages which any social scientist can master in two weeks. Major cost overruns and waste result when the management loses contact with the programmers. This is especially so with large surveys requiring elaborate data files. Research managers must be chosen on their ability to manage all aspects of the project. Documentation on the data base is mandatory and well worth the extra effort. Since "in-house" survey research has a great probability of being interrupted, and should not be stalled during the field work

Chapter 6

DATA PRESENTATION AND STATISTICAL ANALYSIS

6.1 INTRODUCTION

As shown in previous chapters, there are a large number of factors which must be considered when a survey is being planned. The basic research idea, the questionnaire to be constructed, the design of the sample, and the management of all aspects of the survey are crucial to a successful program of survey research. Once the questionnaire has been completed, the necessary editing and quality control procedures employed, the questionnaire answers coded and entered into the computer, the analysis of the results can begin.

Analysis extracts the information contained in the survey responses. All information relative to the research idea, which gave rise to the entire process, is waiting to be released. Too often, the analysis used is either inappropriate to the nature of the data obtained and thus leads to conclusions which are unwarranted, or falls far short of exhausting the potential of the data to provide insight. Care must be exercised in the choice of analysis and substantial thought must be given, at a very early stage, to the nature of the presentation techniques and analysis to be used. It is certainly beyond the scope of this chapter to do more than survey the field, but some basic ideas, philosophies, and techniques can be presented which will begin to reveal the nature of the statistical treatment of the collected data.

To illustrate some of the procedures, the Energy and the Homeowner survey conducted by the Institute for Social and Economic Research will be used. From

TABLE 1

Expenditure on Electricity, Oil and Natural Gas
May 1, 1981 to May 1, 1982
(in dollars)

350	1,600	920	1,305	990	2,500	850	800
1,194	875	550	1,000	1,150	875	970	980
950	931	885	1,075	1,150	800	1,150	1,800
750	570	1,050	538	850	840	1,240	1,874
1,050	640	755	759	1.147	1,000	1,400	1,326
670	1,030	1,200	1,376	1,404	944	700	1,300
1,000	910	1,336	1,000	1,145	1,020	1,060	875
1,750	500	1,060	610	875	1,050	970	821
775	769	550	960	910	660	584	850
1,235	1,980	858	1,300	985	850	850	900
976	897	845	800	1,000	800	800	1,400
1,186	500	950	1,030	780	1,500	1,150	1,120
550	910	640	884	645	790	940	1,350
1,200	710	910	1,650	950	720	900	
1,200	1,240	912	745	1,150	800	960	
1,080	1,640	972	520	780	1,207	850	
1,100	929	550	700	900	784	850	
843	730	755	972	1,040	845	700	
900	625	1,220	103	1,300	954	950	
969	1,200	820	1,350	950	1,400	625	
800	690	870	1,340	1,280	920	1,200	
750	1,075	1,775	698	1,093	1,675	650	
800	700	720	1,140	770	869	1,500	
640	1,138	875	800	760	960	1,300	
762	760	1,020	650	1,170	678	900	

These are "raw" data containing all the information about the annual expenditure on energy by the homeowners. Certainly close examination reveals some information, but, in reality, little can be learned from the data in this form.

To learn more, a summary is desirable. One such summary, termed a frequency table, is often developed. Examination of Table 1 reveals that the smallest expenditure reported was $103, and the largest was $2,500. The frequency table is constructed by sub-dividing the interval from the lowest to the highest data value into a set of distinct sub-intervals or classes. The choice of the number of classes to be created is rather arbitrary, but should be sufficient to be re-

In constructing such a table care was taken that all observed values are accommodated in the set of classes, and each value is placed in one and only one class.

To aid understanding, a graph is sometimes helpful. Derived from the frequency table a histogram is developed as shown in Figure 1.

Figure 1

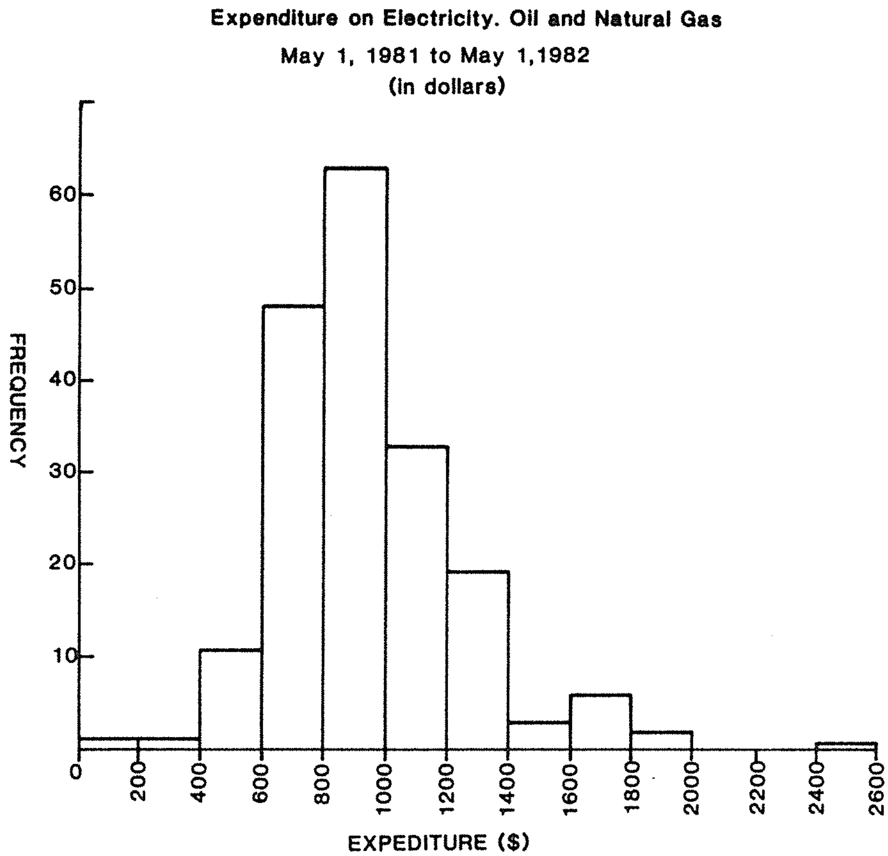


TABLE 3

Stem and Leaf Plot
Expenditure on Electricity, Oil and Natural Gas
May 1, 1981 to May 1, 1982
(in dollars)

$00 Stem	
1	03
2	
3	50
4	
5	50,70,00,00,50,50,50,38,20,84
6	70,40,40,25,90,40,10,98,50,45,60,78,25,50
7	50,75,50,62,69,10,30,00,60,55,20,55,59,45,00,80,80,70,60,90,20,84,00,00
8	43,00,00,75,97,85,58,45,20,70,75,00,84,00,50,75,75,00,40,50,00,00,45,69,50,50,00,50,50,00,75,21,50
9	50,76,00,69,31,10,10,29,20,50,10,12,72,60,72,90,10,85,50,00,50,44,54,20,60,70,70,40,00,60,50,00,80,00
10	50,00,80,30,75,50,60,20,00,75,00,30,00,40,93,00,20,50,60
11	94,86,00,38,40,50,50,47,45,50,70,50,50,20
12	35,00,00,40,00,00,20,80,07,40,00
13	36,05,76,00,50,40,00,00,26,00,50
14	04,00,00,00
15	00,00
16	00,40,50,75
17	50,75
18	00,74
19	80
20	
21	
22	
23	
24	
25	00

This table provides a summary in one table which presents the same frequency information as provided in Table 2, a visual representation similar to that in Figure 1, but with retention of every individual value. In presenting summaries of this nature, a fertile imagination and ingenuity is useful. Space is far too limited here to give examples and to display the range of other possibilities; pie charts, picturegrams of various kinds, multidimensional plots, etc., are but a few which are possible. Typically most software will present a wide menu from which to choose.

In any situation, one must be careful to display the data honestly. Certainly many tricks can, and have, been used to present segments of the data in a manner which might tend to overemphasize the importance of the segment of the data. Manipulation of scales of measurement, suppression of the zero point of the axes on a graph and magnification of segments of the graph, may be misleading. This is not science.

two equal peaks, bimodal and with many peaks or modes, multimodal. An examination of Figure 1 shows that the modal "class" is the class from \$800 to \$1,000.

A third measure of location considers each data point, its value, and position. If the frequency distribution is visualized as being cut out of a flat piece of wood, the arithmetic mean would then be the point of balance and is designated as $\overline{X}$.

If n values in a data set are denoted as X_1, X_2, X_3, ..., X_n, the mean is defined to be,

$$\overline{X} = \frac{X_1 + X_2 + X_3 + .. + X_n}{n}$$

the well known "average" of the data. As a shorthand way of representing this, the summation symbol Σ, is used which designates the sum $X_1 + X_2 + .. + X_n$ as $\sum_{i=1}^{n} X_i$, and the arithmetic mean is written as

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad \text{, or simply} \quad \overline{X} = \frac{\Sigma X_i}{h} .$$

If the data are summarized into a frequency table, the individual values are sacrificed for a tabulation of the number of values falling into various classes. In this situation it is not generally possible to find the median or the mean. However, some of the information can be recovered by making some assumptions about the data values in the classes of the frequency table.

For example, if all those values occurring in a class have the value of the midpoint of the class (clearly not true), an approximate value for the mean is possible. Defining the midpoint of the i^{th} class as X_{0i}, $i = 1,2,..,K$ and correspondingly the class frequency as f_i, $i = 1,2,..,K$, the mean is given by

If the ordered data are divided into ten equal parts, the values which accomplish this partition are called the <u>Deciles</u> denoted D_1 , D_2 ,..., D_9. Between any two consecutive deciles there is exactly 10 percent of the data, and again the median corresponds exactly to the 5th decile, D_5. If the ordered data are divided into 100 equal parts, the values defining the divisions are called the <u>percentiles</u>, P_1, P_2, ..., P_{99}.

6.2.3 Measures of Spread

Another property of general interest is the range of values over which the data are distributed. In examining Figure 1, the spread is clear. If two or more sets of data are compared, often the most striking feature is the relative amounts of concentration or spread exhibited by the data.

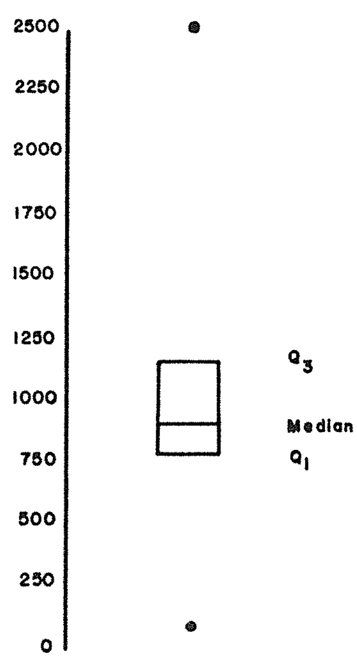
To quantify the spread, various measures are used. In particular, the <u>range</u> R, describes the total length encompassed by the data. For example, the largest energy expenditure for a homeowner was $2,500, the smallest expenditure $103, yielding a range of $2,397.

Data very often exhibit a tendency to be concentrated in the centre of the histogram, and the width of the interval covering the middle 50 percent of the data is given by the difference between the third quartile and the first quartile, Q_3, - Q_1. Similarly $D_9 - D_1$ would measure the length of the interval covering the middle 80 percent of the data. These are quite useful measures of spread.

Another quite different measure of spread involves determining how much the individual data values differ from a measure of the centre. For example, we might consider the difference of all the data values from the mean, $\bar{X}$. One property of such a set of differences is that, no matter what the degree of spread, the differences $(X_i - \bar{X})$ always sum to zero, i.e., It is interesting to note that attempts have been made to summarize these measures of the location and spread in a graphical form. One such presentation is a Box and Dot Plot (one of the names used). The fundamental idea is to show the total range of the data (the extreme values), the centre (the median) and the length covered by the middle 50 percent of the data (the quartiles) in one picture, as shown in Figure 2.

Figure 2

A Box and Dot of Expenditure On Energy Data



6.2.5 Other Data Type Presentations

Nominal data represent a counting of the number of items, or responses, which fall into given classifications. This may seem to be identical with the frequency table encountered earlier, but here the classes do not represent intervals of measurement, but labelled categories of response. For example, in the energy survey, respondents were asked for information concerning the amount, and type of insulation added since the house had been purchased (see Section III, question 6 of the questionnaire). A tabulation of the responses is given in the following table.

TABLE 4

Amount and Type of Insulation Added to Attic

Amount and Type	Number of Responses
None or minimal (saw-dust, horse hair, etc.)	40
4" of fibreglass or cellulose	33
8" of fibreglass or cellulose	46
10" of fibreglass or cellulose	30
More than 10" of fibreglass or cellulose	19
TOTAL	168

As an additional example we note that in Section IV, question 2f the homeowner was asked if payment for their energy conservation steps was made using money from the Canadian Home Insulation Program (CHIP). The responses are classified into two categories, Yes, and No, and are summarized in Table 5.

There is little limit to the extent of the cross tabulations. A common device used involves a two-way table such as presented in Table 6 given according to the response to a third category. A set of two-way tables would be provided, one two-way table for each level of the third category and are referred to as 3-way tables. Multi-way tables provide a means by which important interactions between two or more variables may be detected when a third variable is controlled.

6.2.6 A Bivariate Table for Interval Data

Having described bivariate, or two-way tables for nominal data it is reasonable to expect that similar tables can be provided for ratio, or interval data. Such a table bears a strong resemblance to the frequency table presented in Table 2.

For example, in the Energy Survey, homeowners were asked to provide information on the value of their home, and on their salary level. The responses to these questions are presented in a two-way table, Table 7. What is done is to recode the interval data into discrete categories to simulate nominal responses. There is a loss of information on a scale such as income, but the interaction between that variable and another, here, the value of the home, can be analyzed. Of course other more formal techniques allow such analysis.

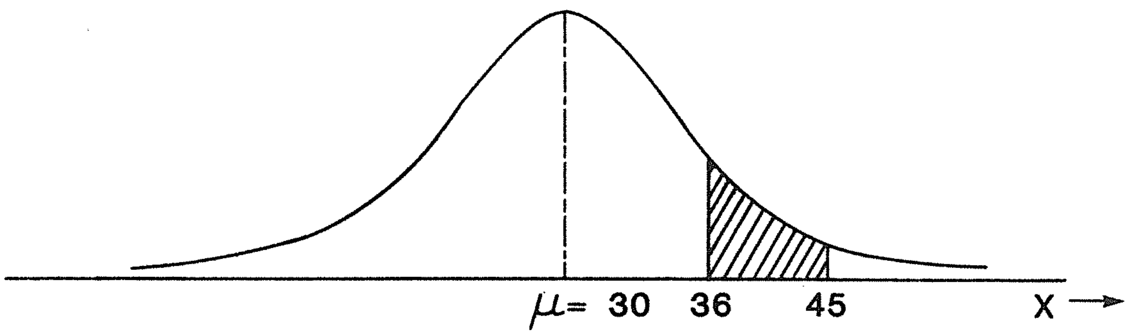
which generalizations are made from a particular set of observed circumstances. Note that if the survey involves the examination of each and every unit in the population, there is no generalization. The reasoning used in such a situation is purely deductive in nature and is common to census analysis.

As part of the analysis, statistical methods differentiate between two distinct but related aspects. The first refers to what is commonly known as estimation, in which, on the basis of sample results, estimates are made of the unknown value of the population characteristics. Assumed is that the sample results come from a population of values which has a mathematical specification (or model), and that this specification involves population characteristics called parameters, the value of which, if known, allow for the complete description of the population model. The task at hand then is to obtain, in the "best" way possible, some indication of the value of these population parameters, based on the observed data from the sample survey.

The second aspect of the statistical methodology is to consider a sample as if it had come from a population with a given description: a mathematical model with specified parameter values. In the light of the observed values in the sample, the question is whether or not the specified population description is reasonable. Or, given two or more samples from populations, whether or not the sample evidence indicates that the populations from which the samples arose are the same. Such situations refer to the Testing of Statistical Hypotheses, or hypothesis testing. Statistical methods provide the criteria by which such decisions concerning the reasonableness of the population model can be made.

Figure 3

A Normal Distribution



The mathematical expression for the normal distribution enables one to determine the probability that the normal random variable takes on a value contained in some interval. For example, if X is a normally distributed random variable with parameter value μ equal to 30, the probability that the random variable X is in the interval from 36 to 45 (i.e., Prob. $(36 \le X < 45)$ is simply an area under the curve and here is represented by the shaded area in Figure 3. To aid in the obtaining of areas, or probabilities like this example, tables of the normal distribution have been prepared and are readily available, as shown reproduced in Table 8.

This table provides areas under the curve associated with a normal random variable with $\mu = 0$ and $\sigma^2 = 1$, which is known as the Standard Normal Probability Distribution. If X is a normally distributed random variable, with known values for μ and σ^2, the standardized normal random variable Z, is obtained from the relationship.

$$Z = \frac{X - \mu}{\sigma} .$$

The sample mean, $\bar{X}$, based on a random sample of size n, is one commonly used statistic whose sampling distribution closely resembles the normal probability distribution, almost regardless of the form of the population from which the sample is selected. In fact, for large n, $\bar{X}$ is approximately normally distributed with mean μ, and variance $\frac{\sigma^2}{n}$, where μ, and σ^2 are the mean and variance of the infinite population from which the sample is selected. This is an example of a statistic which is said to obey the Central Limit Theorem and is a very important result.

To review the main idea, recall that for samples drawn from a population, the sample mean will be randomly distributed. For large numbers of samples drawn from an infinitely sized population the variable $\bar{X}$ will approximate a normal distribution. It is this result which forms the basis for much of the analysis in survey research.

6.3.2 Estimation

The first section of this chapter discussed various presentations of data. If all the values in a population were observable, then the construction of frequency tables, histograms and the like, would be possible. These tables and graphs would show characteristics, similar to those previously discussed, and summary measures could, theoretically, be constructed. These quantities would likely be the parameters of our mathematically described model for the population.

The statistic, $\bar{X}$, being the mean of the random sample from a population, is an estimator of the mean of the population from which the sample was selected. In a particular sample, the value that $\bar{X}$ assumes is an estimate of the population mean, designated μ. This particular value of X from our sample is only an

$$S^2 = \frac{(99.50-947.37)^2(1)+(299.50-947.37)^2(1)+\ldots+(2499.50-947.37)^2(1)}{187}$$

$$= 96,733.4$$

The point estimate of σ^2 from the sample data is \$96,733.4. A point estimate of the population standard deviation, σ is found by taking the square root of the estimate of σ^2, by which we obtain the estimate, S, equal to \$311.02.

From Table 5, nominal data concerning the use of the Canadian Home Insulation Program, an estimate can be obtained of the proportion of homeowners who used CHIP to assist with payment for additional insulation. This estimate is given by the ratio of the number of homeowners who indicated they had used the program to the total number of respondents. In this case, the point estimate is

$$\frac{W}{n} = \frac{66}{168} = 0.39$$

In generalizing this result to the population, extreme care must be taken to ascribe its value to the "correct" sampled population. These respondents were those who had indicated in a previous question that their house had little or no attic insulation when it was purchased by them. The responses in Table 5 are from the population of homeowners who added insulation to the attic in which previously there had been little or none.

Point estimates are informative, and important in any statistical analysis, but it is reasonable to expect that, because only a small part of the population was examined, the estimate based on the sample is not exactly equal to the true population value. Surely it would be surprising if the true mean expenditure on energy by all homeowners in Winnipeg is equal to \$947.37; but the value given by the point estimate is close in value to the true population mean expenditure.

$$947.37 - 1.96 \frac{(311.02)}{\sqrt{188}} \leq \mu \leq 947.37 + 1.96 \frac{(311.02)}{\sqrt{188}}$$

or $\$902.91 \leq \mu \leq \$991.83$.

By changing the degree of confidence, or the Z value, the length of the interval will change. For example a 90 percent confidence interval would use $Z = \pm 1.645$ instead of ± 1.96 as before to get

$$947.37 - 1.645 \frac{(311.02)}{\sqrt{188}} \leq \mu \leq 947.37 + 1.645 \frac{(311.02)}{\sqrt{188}}$$

or $\$910.06 \leq \mu \leq \$984.68$.

Clearly because this interval is somewhat shorter than the previous one, there is less confidence that it does enclose the unknown mean.

A 99 percent confidence interval would use $Z = \pm 2.576$, and would produce the interval estimate,

$$\$888.94 \leq \mu \leq \$1005.80.$$

This interval is of greater length than the other intervals, it hence covers a greater range of possible values and therefore provides a higher degree of confidence that it has covered the location of the true mean expenditure.

The sense in which this is a confidence interval is that, based on probability theory, this outlined procedure will produce an interval which has a specified probability (here 0.90, 0.95 or 0.99) of covering, or enclosing the true mean, μ. As a result, when executing a survey obtaining the results, and using this interval estimation procedure, the particular interval is but one of many which could have been obtained. It is reasonable to expect that the particular interval obtained will be one of those which does enclose μ. Thus, there is a specified level of confidence that μ will be within the stated interval.

For many sample statistics used as estimators of a population parameter, it is possible to obtain such a sampling distribution. When this can be done, it an hypothesized probability distribution for a variable in the population, or some other condition hypothesized for the population. This hypothesized condition is formulated into a precise statement and is labelled as the _null hypothesis_, designated as Ho. This null hypothesis is the condition that is to be tested with respect to the observation made in the sample. To provide a basis for comparison a second hypothesis, an _alternative_ to the null hypothesis also is stated and it is designated as Ha.

Clearly, the statement provided in the null hypothesis for the population either is true, or not. On the basis of what is observed in the sample, the null hypothesis is either rejected, or it is not rejected.

Within the framework of the "classical" hypothesis testing procedure, there is the "true" state of nature - either the null hypothesis is true, or it is not true, and the correct state is unknown. There is an action or decision to be made - either the null hypothesis is to be rejected, or it is not to be rejected.

A schematic arrangement of these two aspects shows the results of the various possibilities, and the decision taken.

It is common to set a null hypothesis which is discordant with the theory on the conjecture of the researcher. For example, in testing the theory, that there is a differential between male and female wages, one might create a null hypothesis which explicitly stated that the mean wage of men and women are equal (i.e., Ho, $\mu_M = \mu_F$). In this way rejecting the null hypothesis indirectly supports the theory which purports to explain wage differentials.

Once a rejection rule has been established, it is then possible to evaluate the probabilities of the two types of errors previously defined. We can find the probability of rejecting the null hypothesis when it is in fact true, and we can find the probability of not rejecting the null hypothesis when it is not true (and therefore the alternative condition is true). These two probabilities are denoted by α and β respectively. Thus,

Pr { Rejection of Ho, when the Ho is true } = α

and

Pr { Failure to reject Ho when the Ho is false } = β

where Pr { is read as "the probability that".

A complementary event of rejecting the null hypothesis when it is false, being the making of a highly desirable correct decision is of particular importance. The probability of making this particular correct decision is found to be $1 - \beta$, and is referred to as the "Power" of the test.

Given a null hypothesis and its associated alternative hypothesis, a rejection rule is established. For a specified rule there will be a particular value to α and to β. If a different rejection rule were considered, it too would have a value for the probabilities of Type I, and Type II error which in all likelihood would be different. Ideally, we would like to find a rule for which the resulting values for α, and β are very small.

not sufficient to cause us to doubt the null hypothesis, and hence the null hypothesis is not rejected. This rejection region is called the significance region, or the critical region. When the observed result falls into the rejection region, the null hypothesis is rejected, and the observed result is said to be "statistically significant".

In recent years, the trend in statistical methodology has been to move away from this rather rigid rejection rule format in hypothesis testing. This movement has been due in no small part to the wide acceptance of values like 0.01, and 0.05 for α without much question. There is nothing magic about such levels, and in one situation they might well be smaller than is really necessary, and in others a value of α at 0.01 may be too large.

Lately attention has been focused on evaluating the observed results in comparison with the null hypothesis and obtaining a measure of how unusual the observed results are under the stated situation in the Ho. The probability of observing a sample result as much in disagreement with the null hypothesis, or more so in the direction indicated in the alternative hypothesis, when the null hypothesis is true, is evaluated. This probability is called the P-value, or the observed significance level of the sample result. This P-value is then reported.

Clearly, if the P-value is small, that is evidence against the null hypothesis, whereas a large P-value indicates that under the null hypothesis the observed situation is not unusual. The prime advantage of simply reporting the P-value of an observed result is that much more information is being provided to the reader than is done under a simple statement that "the null hypothesis was rejected". Having seen the reported P-value, the reader is free to decide how the result is to be regarded and to decide what should be regarded as large or as small.

TABLE 10

Statistical Analysis System
Table of Diffdeg by Auto

DIFFDEG AUTO

FREQUENCY PERCENT ROW PCT COL PCT	0	1	TOTAL
.	7 . . .	13 . . .	. .
-10	1 0.30 100.00 0.65	0 0.00 0.00 0.00	1 0.30
-8	1 0.30 100.00 0.65	0 0.00 0.00 0.00	1 0.30
-5	1 0.30 33.33 0.65	2 0.59 66.67 1.09	3 0.89
-4	1 0.30 50.00 0.65	1 0.30 50.00 0.54	2 0.59
-3	6 1.78 100.00 3.92	0 0.00 0.00 0.00	6 1.78
-2	0 0.00 0.00 0.00	7 2.08 100.00 3.80	7 2.08
0	44 13.06 44.44 28.76	55 16.32 55.56 29.89	99 29.38
TOTAL	153 45.40	184 54.60	337 100.00

DIFFDEG AUTO

FREQUENCY PERCENT ROW PCT COL PCT	0	1	TOTAL
1	1 0.30 20.00 0.65	4 1.19 80.00 2.17	5 1.48
2	16 4.75 39.02 10.46	25 7.42 60.98 13.59	41 12.17
3	17 5.04 53.13 11.11	15 4.45 46.88 8.15	32 9.50
4	29 8.61 55.77 18.95	23 6.82 44.23 12.50	52 15.43
5	18 5.34 47.37 11.76	20 5.93 52.63 10.87	38 11.28
6	2 0.59 33.33 1.31	4 1.19 66.67 2.17	6 1.78
7	4 1.19 22.22 2.61	14 4.15 77.78 7.61	18 5.34
8	5 1.48 45.45 3.27	6 1.78 54.55 3.26	11 3.26
TOTAL	153 45.40	184 54.60	337 100.00

DIFFDEG AUTO

FREQUENCY PERCENT ROW PCT COL PCT	0	1	TOTAL
9	0 0.00 0.00 0.00	2 0.59 100.00 1.09	2 0.59
10	4 1.19 57.14 2.61	3 0.89 42.86 1.63	7 2.08
11	1 0.30 100.00 0.65	0 0.00 0.00 0.00	1 0.30
12	1 0.30 50.00 0.65	1 0.30 50.00 0.54	2 0.59
13	0 0.00 0.00 0.00	1 0.30 100.00 0.54	1 0.30
15	1 0.30 50.00 0.65	1 0.30 50.00 0.54	2 0.59
TOTAL	153 45.40	184 54.60	337 100.00

TTEST PROCEDURE

VARIABLE: DIFFDEG

AUTO	N	MEAN	STD DEV	STD ERROR	MINIMUM	MAXIMUM	VARIANCES	T	DF	PROB > |T|
0	153	2.63398693	3.44840091	0.27878669	-10.00000000	15.00000000	UNEQUAL	-0.5370	316.4	0.5917
1	184	2.83152174	3.25574496	0.24001670	-5.00000000	15.00000000	EQUAL	-0.5398	335.0	0.5897

FOR HO: VARIANCES ARE EQUAL, F'= 1.12 WITH 152 AND 183 DF PROB > F'= 0.4558

$$P\ (Z \geq .537).$$

This probability can be found from a table of the standard normal probability distribution which was given in Table 8. This probability is approximately 0.295, and we can report that the P-value or significance level of our observed mean temperature difference between homes with, and without automatic thermostats is 0.295.

With a P-value this large it would seem that the observed difference in the survey is not sufficiently large for us to conclude that homeowners with automatic thermostats change the temperature of their home more than homeowners with manual thermostats. Certainly we observe a difference in the survey, but such a difference does not seem to be unusually large.

It should be noted in connection with the example presented herein, and the output from the computer that there are one or two slight differences. The computer output refers to a T-test procedure and presents a value for Prob > |T|. With the sample sizes involved here, there is essentially no difference between the procedure outlined here and that shown in the computer output.

Also, the computer shows a value for T equal to -.5370 whereas we have shown a value of $Z = .537$. As mentioned above the Z and the T are essentially the same here, and the difference in sign on the value of the statistic is due to the computer taking $\overline{X}_M - \overline{X}_A$.

The computer program, however has carried out a test procedure which has used as its alternative hypothesis that $\mu_A \neq \mu_M$, simply that the mean differences are not equal. Our procedure has implied a particular direction to the alternative, that $\mu_A > \mu_M$. The P-value found by the computer, which is appropriate for its alternative hypothesis is exactly (or almost) twice as large as for our directional alternative. This is exactly what one should expect to happen for these two alternative hypotheses.

6.3.5 Regression and Correlation

As previously noted, data are often obtained not just as single values, but as pairs of values. Such data can be summarized as shown in bivariate, or two-way tables as, for example, Table 7 which illustrates the information provided on house value and income for homeowner respondents. Of interest may be the nature of the relationship between the two variables.

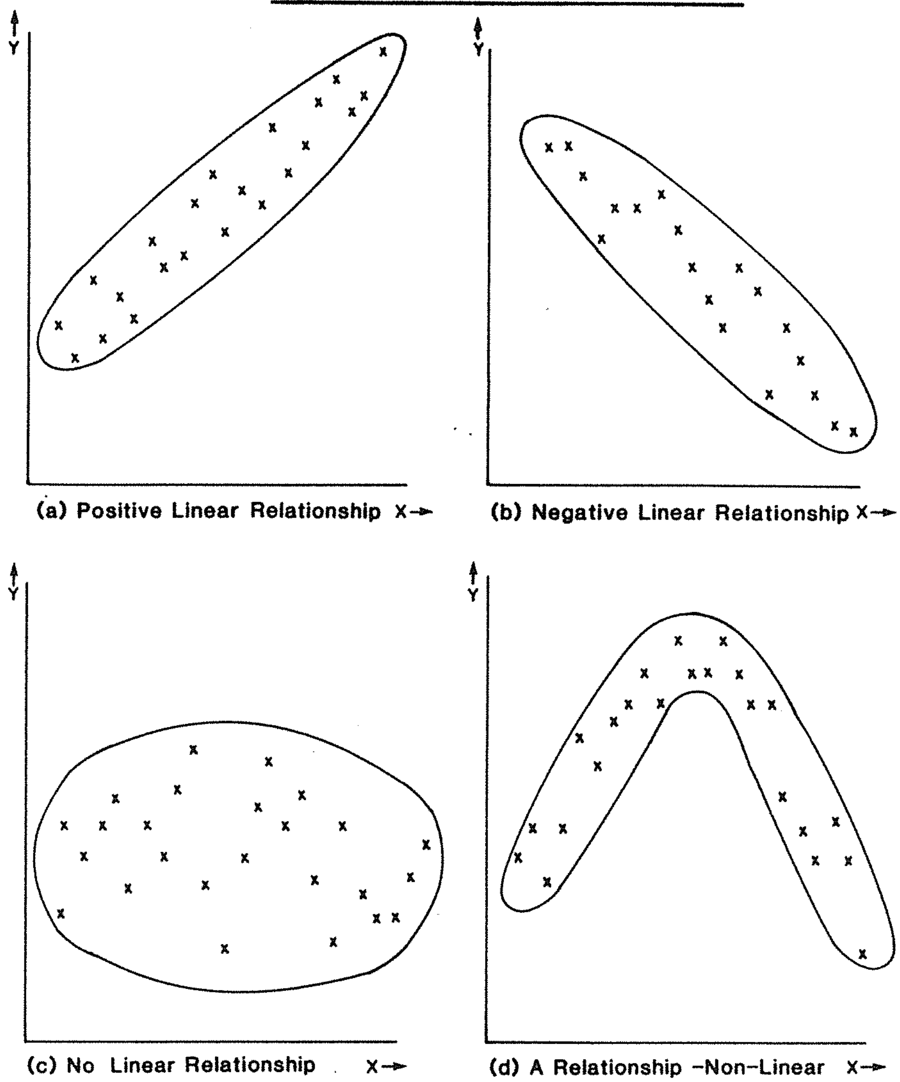
For the data in the example of Table 7, it is not surprising if homeowners with high incomes tended to have houses in the higher price range, and vice-versa. An examination of Table 7 does seem to suggest exactly that, as there are very few entries in the upper right and in the lower left portions of the table, but a heavy concentration of responses in a band along the diagonal in the table from the upper left corner to the lower right corner.

It is possible to construct a picture of these individual values, by plotting each pair of values (income, house value) on a graph. Such a plot is referred to as a scatter diagram or a scattergram.

There are a host of possible patterns which might become evident in a scatter diagram. To illustrate, Figure 4 below shows several idealized scatter diagram patterns. In (a) there is evidently a tendency for one value to increase as the second value increases. This particular pattern suggests a linear relationship and the upper slope of the points indicate the linear relationship is positive. The scatter diagram (b) is an illustration of a negative linear relationship in that the variable labelled Y decreases as the X variable increases. Plot (c) shows no particular relationship, there is general scattering of the points inside the prescribed figure (almost a circle). In (d) there clearly is a relationship, as X increases in value we see an increase in the Y variable, until a maximum is reached, and then a decrease. The relationship shown here however is certainly not linear; perhaps it is quadratic.

Figure 4

SCATTER DIAGRAMS OF IDEALIZED DATA



Using this idea of the linear relationship representing the points in our data set, each observed X value, determines a corresponding Y value from the equation of the straight line. This Y value is not, of course, the value observed in the data set, but is a value which is "predicted" by the equation of the straight line. In order not to confuse this predicted value, from the actuis a minimum. The procedure being used here is called <u>Least Squares</u>.

Using this principle, the equation of the "best fitting" straight line to the n pairs of values (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n) is

$$\hat{Y}_i = a + bX_i, \quad i = 1,2, \ldots, n$$

where $$b = \frac{\Sigma(X_i - \overline{X})(Y_i - \overline{Y})}{\Sigma(X_i - \overline{X})^2}$$

$$a = \overline{Y} - b\overline{X},$$

and $\overline{X}$ and $\overline{Y}$ are the sample means of the X and Y variables respectively.

There are, of course, variations on the form of the slope b. Using some algebra, it can be shown that

$$b = \frac{n\,\Sigma X_i Y_i - (\Sigma X_i)(\Sigma Y_i)}{n\,\Sigma X_i^2 - (\Sigma X_i)^2}$$

The procedure that has been outlined here can be used to find the equation of a straight line through any data points, whether or not a linear relationship is suggested by a scatterdiagram. It is a mathematical technique which finds the "best" line, even though that line may not be very good, in terms of how well it represents the data points.

From a statistical point of view, consider that the pairs of values, X and Y, have a particular structure. Assume that the X variable is a fixed variable called the <u>independent variable</u>, and Y is a random variable called the <u>dependent variable</u>. The straight line fitted to the pairs is then a measure of the linear dependence relationship of the dependent variable, Y, on the independent variable X.

Because of the normality of a and b, it is possible to obtain estimates for α and β, and to convert these point estimates into confidence interval estimates for α and β in a way completely analogous to that described for the population mean μ. In addition, the normality of the estimators a, and b, permits a test of hypotheses about α and β. Finally, the estimated linear relationship found enables a prediction of values for the Y variable at specified values of X, which perhaps were not included in our responses, and to measure how precise these predictions are through a confidence interval, or prediction interval approach. This is the basis for forecasting.

Of particular interest to us, is a test for the slope parameter β in the model. If β has, in fact, the value zero, the model indicates that X and Y are not linearly related, because under this condition the model is written

$$Y_i = \alpha + \varepsilon_i, \quad i = 1,2, \ldots, n$$

which does not involve X at all.

In a given situation in which a value for the estimate of the parameter β has been obtained, it is possible to construct an hypotheses test to determine if a linear dependence relationship does exist between X and Y. We establish a null hypothesis that β equals zero,

$$Ho: \beta = 0$$

against an appropriate alternative hypothesis, for example

$$Ha: \beta \neq 0$$

Under this null hypothesis, and given the conditions specified for the linear model, the quantity,

$$\frac{b - \beta}{S_{y.x} / \sqrt{\Sigma(X-\overline{X})^2}}$$

where $S_{y.x} = \sqrt{\frac{\Sigma(Y-\hat{Y})^2}{n-2}}$ is a random variable which has a probability where, as before, the data are interval and obtained as n pairs of values (X_i, Y_i), i = 1,2, ..., n.

It is important to note that the correlation coefficient, r, measures the linear association between X and Y. A value of +1 or -1 indicates a perfect linear association, and the value of zero indicates no linear association. A value of r close to zero does not indicate that there is no relationship between X and Y, only that there is no strong evidence that the relationship is linear. In addition, a value of r close to +1 or -1 is an indication of a linear interdependence relation, but it is most certainly not an indication of any cause and effect relationship.

Under appropriate circumstances and assumptions, r can be taken to be an estimate of a corresponding population characteristic ρ, the population correlation coefficient. The quantity ρ is a parameter of a probability distribution for X and Y jointly, known as the bivariate normal probability distribution. If X and Y are bivariate normal random variables, a particularly interesting situation arises when the true value of ρ is zero, in which X and Y are independent random variables; i.e., there is no relation between X and Y.

In a data set, representing the results of observing variables X and Y from a bivariate normal distribution, we may obtain a value for the correlation coefficient estimator r, and ask whether or not its value is such that it could have arisen from a bivariate normal population in which the true value for ρ is zero. Under the sampling procedure and assumptions, the statistic

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

is appropriate for testing a null hypothesis

$$Ho: \rho = o$$

against an appropriate alternative, i.e., Ha: $\rho > 0$.

This statistic follows a Student's t-probability distribution and

The P-value for this value of r, and hence t, is extremely small, less than .001. There is here, a statistically significant linear association between house value and homeowner salary level.

6.3.7 Contingency Tables and Tests of Independence

Very often in survey research, the type of data obtained is generated by counting, or enumerating the number of occurrences of some event, or response. In the Energy Survey, enumerated are the number of homeowners who obtained a CHIP grant, and the number of homeowners who added 4 inches of fibreglass or cellulose insulation to their attic. Such data has been considered previously and examples were displayed in Tables 4, 5 and 6.

For the purpose of illustrating a slightly larger table of a type similar to Table 6, Table 11 presents the results of two questions asked in the survey. One referred to the addition of insulation to the attic by the homeowner; the other concerned the education level of the homeowner. The results of a cross tabulation of these into a two-way table is presented in Table 11.

6.3.8 Independence

Consider the nature of a relationship which can be examined for contingency tables. The approach used will be rather general and symbolic but reference to Tables 6 and 11 should provide clarity to the ideas presented abstractly.

Suppose that a set of data is classified according to two criteria; the first has two levels labelled 1 and 2; and the second has three levels, A, B, and C. Tables 12 (a), (b), and (c) display the arrangement of the cell counts in a 2 x 3 contingency table, and in the marginal tables for each criterion separately.

TABLE 12

Two-Way Marginal Tables For Enumeration Data

(a)

Criterion 1	Criterion 2: A	B	C	Totals
1	N_{1A}	N_{1B}	N_{1C}	N_1
2	N_{2A}	N_{2B}	N_{2C}	N_2
Totals	N_A	N_B	N_C	N

(b) Criterion 1

Level	Number
1	N_1
2	N_2
Totals	N

(c) Criterion 2

Level	Number
A	N_A
B	N_B
C	N_C
Total	N

TABLE 13

Cell and Marginal Probabilities

	Criterion 2			
Criterion 1	A	B	C	Total
1	P_{1A}	P_{1B}	P_{1C}	P_1
2	P_{2A}	P_{2B}	P_{2C}	P_2
Total	P_A	P_B	P_C	1

If it should happen that the value of these probabilities are known, we can determine, for our survey in which there are N respondents, how many counts to expect in the various cells. Expected is $N \cdot P_{1A}$ in the 1A cell, $N \cdot P_{2A}$ in the 2A cell, etc. It is important to note that we likely will not obtain exactly this "expected" number of counts when the survey results are obtained, but the observed values will be similar.

The relationship of interest is directly related to the probabilities. Of particular interest is how the marginal probabilities P_1 and P_2 for Criterion 1, and P_A, P_B, and P_C of Criterion 2 relate to the probabilities in the individual cells of the table.

In probability theory, two events, call them R, and S, are said to be <u>independent events</u> if the probability of them occurring together is equal to the product of the probabilities of their occurring separately. We express this relationship as,

$$P\ (R \text{ and } S) = P(R) \cdot P(S)$$

if, however, $P(R \text{ and } S) \neq P(r) \cdot P(S)$

event R and event S are called

Designate the count observed in the actual survey for cell 1A as O_{1A}, for cell 1B as O_{1B} and so on for all cells. If E_{1A}, the "expected" number under independence is similar to O_{1A}, E_{1B} similar to O_{1B} etc., this agreement lends support to the idea that the Criteria are independent as hypothesized. However, if one or more large differences are noted between the E's and the O's, this suggests that the condition of independence used in finding the expected numbers was not true and the null hypothesis should be rejected.

In determining the degree of agreement or disagreement between the observed and expected numbers computed is the statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the observed number in the i^{th} cell, E_i is the expected number in the i^{th} cell, and k is the number of cells in the table. This statistic when computed under the null hypothesis of independence in a contingency table with r rows, and c columns is a random variable which has a probability distribution approximately like the Chi-square probability distribution with (r-1)(c-1) <u>degrees of freedom</u>.

The "degrees of freedom" here is the parameter of the Chi-square probability distribution, and its value determines the shape of the curve representing the form of the distribution. As for the Normal distribution, and the t-distribution, the Chi-square probability distribution has been tabulated and table showing the various values of χ^2, the Chi-square random variable, for specified degrees of freedom, and for selected probability levels is given as Table 14.

This table allows completion of the test of hypothesis for the independence in the contingency table by providing the critical values of the Chi-square random variable at a specified level of significance. The observed value of χ_2 from the test statistic exceeds the critical value from the table, the null hy-

The only issue left to be addressed is how to obtain the marginal probabilities P_1, P_2, P_A, P_B, and P_C used in funding the expected values? Of course, they are never observed directly but the survey results produce good estimates of their values. From tables 12(b) and 12(c), N_A/N is an estimate of P_A, N_1/N is an estimate of P_1 and so on. It is these estimates of the true probabilities that are used in finding the expected number in each cell, under the condition of independence. Thus,

$$E_{1A} = N \cdot \frac{N_1}{N} \cdot \frac{N_A}{N}$$

$$= \frac{N_1 \cdot N_A}{N}$$

$$E_{2B} = \frac{N_2 N_B}{N}$$

and similarly for each cell.

Now, consider the actual examples obtained from the Energy Survey. Are the two categories, the addition of various amounts of insulation to the attic, and the educational level of the homeowner, independent?

Given the null hypothesis;

Ho: Added Insulation and Education are independent

and the alternative,

Ha: Added Insulation and Education are not independent.

Under this null hypothesis, the expected cell frequencies can be determined. For example, the expected count in the cell associated with "none or minimal" insulation added, and "at most High School" education level is found from the corresponding row and column totals in Table 11 as

$$\frac{(36)\ (100)}{157} = 22.9$$

This χ_2 statistic is approximately distributed as a Chi-square random variable with (4)(2) = 8 degrees of freedom. At the .05 level of significance, and with 8 degrees of freedom, Table 14 provides the critical value 15.51. Our observed χ^2 value of 5.341 does not exceed 15.51 and hence we are unable to reject the null hypothesis. It thus appears that, from these data, there is insufficient evidence to conclude that level of education has an influence on the addition of insulation to the attic.

One note of caution must be sounded at this stage. The Chi-square probability distribution is only an approximation to the probability distribution of our statistic

$$\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i}$$

The approximation is a good one, providing accurate probability level assessment if the <u>expected numbers</u> in each cell of the contingency table are greater than or equal to 5. As the computer has warned us here, such is not the case and we must treat our results with care. With the very large difference between the value of the test statistic (5.341) and the critical value (15.51) we likely are quite justified in not rejecting the null hypothesis in spite of the failure of the assumption on the expected values. The failure of the Chi-square here is an excellent example of a sample which has failed to generate sufficient information to undertake detailed analysis. While samples to test very simple phenomena are surprisingly small as social phenomena become more complex and response categories to questions increase, there may be diminished ability to undertake sophisticated analysis. One option might be to recode the response categories and collapse the cells. This reduces the information content in the variables, but may permit further analysis.

The computer output has also indicated that the P-value, or observed level of significance for our observed χ^2 value is 0.0001. Obviously, a χ^2 value as high as 25.792 or higher under a condition of independence is extremely unlikely to occur. It did occur here however, and hence the condition of independence is highly suspect.

This section of the chapter has present a condensed treatment of classical statistical methods as they are commonly applied to survey research. This certainly does not exhaust the tests which may be applied. Many researchers are enamoured with "nonparametric" statistics and the analysis brims with Mann-Whitney, Kendall's tau and Gamma tests (see Table 16). Before concluding, it is useful to consider procedure on appraising sample quality and how a final report should be presented.

6.4 ANALYSIS OF NON-RESPONSE

No survey will ever produce 100 percent coverage, or will produce questionnaires in which every question has been completed. Non-response is a fact of life in survey research, and increasingly so as individuals become more protective of privacy. The survey researcher must be prepared to deal with imperfect and partial samples. Some respondents refuse to answer any questions (refusal non-response) while others will be selective about questions (item non-response). The key to coping with non-response is to obtain information which allows comparison of refusers and respondents. For this reason, sample frames derived from administrative data sets (property tax records, motor vehicle registrations, association memberships and the like) are desirable as they often provide background data on each member of the sample which can be incorporated into the analysis and data base prior to actual interviewing.

Although a result which indicates that the two groups (responders and recontacted refusers) are not significantly different may be cause for some relief, it is important to stress that these tests pertain only to those variables which were used in the recontact; these will be but a small subset of all the variables used in the survey, and circumspection is essential in proclaiming that non-response is not a problem.

What happens if there are significant differences in the two groups? All is not lost, although the power of the tests and conclusions certainly are tempered. For example, if responders and refusers are significantly different with respect to homeownership, and the point of the survey is to discover political preferences, one might sub-divide the sample of responders into those who do and do not own a home. The political attitudes for each group may then be evaluated. If there is a significant difference, then one could infer that since there is a relatively high rate of refusal, the conclusions with respect to political attitudes may be biased. Clearly this assumes that homeownership is correlated with political attitudes; in fact it probably is a proxy for a complex set of other socio-economic attributes. On the other hand, if homeownership in the set of responders appears not to be significantly related to political preference, there is some basis for believing that refusals have not upset the validity of the survey, at least with respect to questions pertaining to political preference.

It is patently obvious that these procedures are quite conjectural. There is no formal statistical procedure which will alleviate the uncertainty produced by low response rates. While these "partitioning" and recontacting techniques do permit one to recover some lost ground produced by a poor response, there are limits to application of such analysis. A response rate of 20 percent must be perhaps even to provide some discussion on their direction and magnitude moves beyond what appears to be standard practice, certainly in the applied policy arena.

An alternative does exist provided the researcher is prepared to make some assumptions about the structure of the data, especially about the underlying mathematical function of the variable with impaired observations. For example, if the income field is marked by item non-response, yet this variable figures prominently in the analysis (as it usually does), one can estimate income as a function of socio-economic variables. The observations for respondents are separated, and a predictive model is constructed wherein income (in this case income) is made a function of others such as education, age, gender etc. The usual approach is to employ regression. The estimated equation is then used to "predict" values for the missing data by inserting information provided by the item non-responders on education, age and gender. The regression approach is used because, by assuming that errors are normally distributed, and if a "reasonable" model can be constructed to predict the missing values (and reasonable will not be defined here), the insertion of estimated values for the missing data will generally not bias the results of other tests undertaken in the analysis. The real virtue of this approach is that it allows one to incorporate as much information as possible into statistical tests, without running a great danger of biasing these tests. Typically, the power of tests is raised such as, for example, increasing the cell counts in a Chi-Square test. Of course this approach does rest on the assumption of normality in the error structure of the data, the basic assumption of parametric procedures, but this is usually a reasonably safe approach.

times. Indeed social science generally does not undertake enough replication, and surprisingly seems to assume that survey research is much more generalizable than, say experiments in physics. The history of science is littered with charlatans who have hoodwinked the profession; replication is an essential safeguard against this retrograde practice.

Second, the theory underlying the research should be explained. In relatively straightforward survey research, such as opinion polling by telephone, the explicit theory may actually be little more than natural curiosity. For policy research and development of an academic discipline, there will be theoretical bases for the research which will generate clear hypotheses. While these should appear "up-front" and purists may argue that hypotheses should not be altered in the face of statistical tests which suggest alternative hypotheses, in fact, much research proceeds recursively, and theoretical sections and hypotheses are frequently inserted after the statistical analysis is complete. This reflects the bias in social science and policy analysis, wherein weak statistical tests are more often taken to be evidence of poor research design, rather than a manifestation of reality. Survey research which produces data that do not support significant relationships, (i.e., clear rejection of null hypotheses) may in fact be indicating something valid about reality, and need not be evidence of poor design.

Third, once the hypotheses have been stated, it is crucial that the variables contained in these hypotheses be operationalized. In the final report the researcher ought to make explicit, if that is not apparent from the discussion, how the questions included can be expected to produce usable variables for testing. Commonly this is the weakest part of final reports (and grant applications). In some cases, the theory speaks very clearly to the variables needed

Finally the substantive conclusions of the survey need summary. These should relate back to the main hypotheses, and should be unequivocal and accurate. If the hypotheses can not be rejected (i.e., the theory under test fails to receive strong support in the context of the survey) then this must be reported clearly. Also, do not overstate the results.

A number of appendices should appear in any final report. These include a complete copy of the questionnaire, all one-way frequency tables for each question, and a brief description of field operations, especially if unusual procedures were used, or unusual circumstances were encountered. It is also valuable to include an analysis of non-responses and evaluation of sample quality. Finally, the associated data files generated by the survey should be documented, although many prefer to make this a separate document.

There is an obvious tendency to misreport poor surveys, or one's which fail to achieve acceptable responses. Although in the interest of science all research must be completely documented, much survey research is quite poorly reported. If a report does not permit an accurate appraisal of all elements (sample frame, questionnaire, testing, etc.) then it must be assumed that the survey went poorly unless proved otherwise.

GLOSSARY

Auxiliary Information

In some cases, known population parameters can be used to provide auxiliary information from which sample estimators are adjusted.

Bias of an Estimator

The difference between the sample attribute (e.g., average) and the corresponding population parameter.

Causation

The unambigious demonstration of action and reactions. The usual requirements for an event A to cause an event B, are:

1. A and B must always be present together.
2. A must precede B.
3. Non-spuriousness (i.e., other explanations of B are not available).

Despite the apparent obviousness of these three "rules" this question remains incompletely resolved and one of the more important areas of philosophical study.

Closed Questions

Questions where the respondent is forced to select from a fixed list of response values.

Cluster Sampling

Sampling undertaken when the population occurs naturally in groups (e.g., city blocks). Clusters are randomly selected, and random selection of units may proceed with the selected clusters. This is not to be confused with stratified

Cross-Sectional Surveys (Data)

This is information all at a single point in time. Often different regions, cities or individuals are used to provide the variation for a behavioural analysis.

Dependent Variable(s)

Changes in the value of independent variable(s) cause changes in the dependent variable(s). Usually, there is only one dependent variable.

Double-Barreled Questions

Questions which actually contain two or more embedded questions.

Efficiency

The standard error of an estimator.

Efficient Estimator

An efficient estimator is one where the distribution of the sample attribute (mean, variance, etc.) for all possible samples has low variance.

Estimator

The sample attribute from which population parameters are inferred. For example, from the sample average (mean) we infer or estimate the population average (mean).

Haphazard Sampling

The units for the sample are chosen because of the convenience of their inclusion.

<u>Levels of Measurement</u>

All variables can be measured at various levels. (Nominal, Ordinal, Interval and Ratio Measurement).

<u>Longitudinal (Time Series) Surveys or Data</u>

These surveys examine an individual or set of respondents over time.

<u>Measurement Ratio (Scale)</u>

This scale is the lowest measurement scale and variables in this scale can be handled with the full range of algebraic operations.

<u>Multi-Phase Sampling</u>

Here a sample is selected from the population and surveyed. Then, more detailed data can be obtained by resampling the sample for a smaller group.

<u>Multi-Stage Sampling</u>

Here several levels of clusters (or strata) may be identified. At stage one for example, cities may be selected, at stage two census tracts, at stage three enumeration areas, etc.

<u>Nominal Measurement (Scale)</u>

This is the highest level of measurement and consists merely of a labeling scheme to distinguish between values of a variable (e.g., gender).

<u>Non-Sampling Error</u>

Errors in estimating population parameters from sample attributes which are induced by the mechanics of the survey (e.g., refusal).

<u>Open Questions</u>

A question which allows the respondent to control the response values.

Probes

In a face-to-face or phone interview, probes such as "I am not completely clear about that, do you mean....?" are useful. Care must be taken to make probes neutral.

Pseudo Random Number Generator

A computer algorithm which produces random numbers.

Question Type

Attitude	-	What people say they want
Beliefs	-	What people think is true
Behaviour	-	What people do
Attributes	-	What people are

Questionnaire

These are completed by the respondent. Although a certain confusion can exist since one designs a "questionnaire" for interviews as well. Also, for some forms of questionnaire surveys such as the mailout, a certain small amount of interaction is possible between the respondent and researcher.

Quota Sampling

Certain groups are included on a quota basis in an attempt to make the sample representative.

Random Sample

A sample chosen according to a set of probability rules.

Refusal Non-Response

Refusal to participate in any aspect of the survey.

Self-Report Data

This is information supplied by the respondent and not corroborated from other secondary sources.

Simple Random Sampling With Replacement (SRSWR)

Population members are allowed to reappear any number of times in a sample and are selected such that each appearance in the sample has equal probability.

Simple Random Sampling Without Replacement (SRSWOR)

Each member of the population has an equal chance of appearing in the sample only once.

Skip Pattern (Logic)

Many surveys use a series of screen questions to guide respondents around sets of questions. As the number of skips is increased, so too are interviewing and data entry costs. Complex skip patterns are usually not possible in mailout surveys and must always be handled with care.

Standard Deviation

The square root of the variance. When referring to a sample, it is also known as the standard error.

Stratified Random Sampling

The population is divided into sub-groups (where each sub-group is as homogeneous as possible). These sub-groups or strata are then each sampled according to some probability rule.

FOR FURTHER READING

The following are just a few of the many books, monographs and articles in the area of survey research.

Bailey, K.D. Methods of Social Research. New York: The Free Press, 1978.

Bradman, N.M. and S. Sudman. Improving Interview Methods and Questionnaire Design. San Francisco: Jossey-Brass, 1979.

Dillman, D.A. Mail and Telephone Surveys: The Total Design Method. New York: Wiley, 1978.

Maclean, M. and H. Genn. Methodological Issues in Social Surveys, Atlantic Highland. Humanities Press, 1978.

Rosander, A.C. Case Studies in Sample Design. New York: M. Dekker, 1977.

Sudman, S. Applied Sampling. New York: Academic Press, 1976.

BIBLIOGRAPHY

Abt, Clark C., ed. _Perspectives on the Costs and Benefits of Applied Research._ Cambridge, Mass: Abt Books, 1979.

Anderson, A., A. Basilevsky and D. Hum. "Measurement: Theory and Techniques". Ch.7 in Rossi _et al_, _Handbook of Survey Research_. New York: Academic Press, 1983.

Anderson, A., A. Basilevsky and D. Hum. "Missing Data Problems in Survey Research". Ch.9 in Rossi _et al_, _Handbook of Survey Research_. New York: Academic Press, 1983.

Babbie, Earl. _Survey Research Methods_. Belmont, California: Wadsworth Publishing Co. Inc., 1973.

Backstrom, C. and G. Mursch. _Survey Research_. Evanston, Ill.: Northwestern University Press, 1963.

Barnett, V. _Elements of Sampling Theory_. The English Universities Press Ltd., 1974.

Bradburn, N.M., S. Sudman and Associates. _Improving Interview Method and Questionnaire Design: Response Effects to Threatening Questions in Survey Research_. San Francisco: Jossey-Bass, 1979.

Cochran, W.G. _Sampling Techniques_. (Third Ed.). New York: John Wiley & Sons, 1977.

Deming, W.E. _Some Theory of Sampling_. New York: Dover Publications, Inc., 1950.

DeNeafville, Judith I. _Social Indicators and Social Policy_. New York: Elsevier, 1975.

Dillman, D.A. _Mail and Telephone Surveys: The Total Design Method_. New York: John Wiley and Sons, 1978.

Etzioni, Amatai, ed. _Policy Research_. Leiden: Brill, 1978.

Glock, Charles, ed. _Survey Research in the Social Sciences_. New York: Russell Sage Foundation, 1967.

Hansen, M.H., W.N. Hurwitz and W.G. Madow. _Sample Survey Methods and Theory, Volume I_. New York: John Wiley & Sons, 1953.

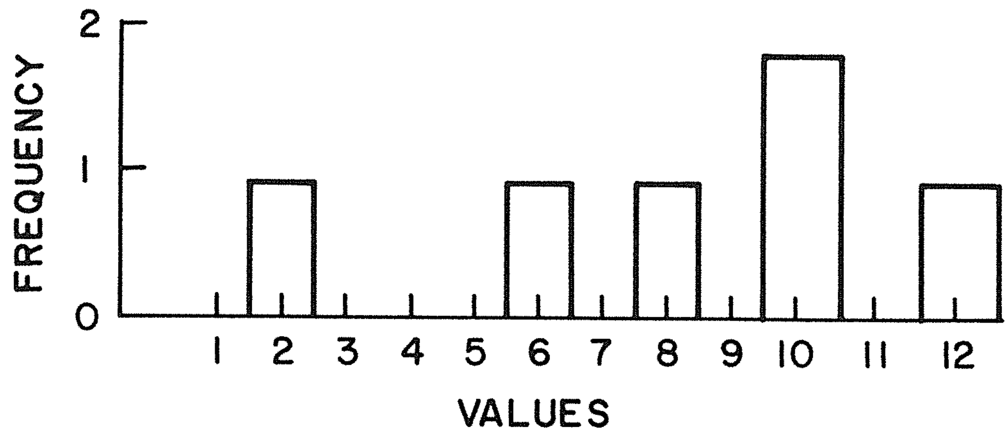
Hansen, M.H., W.N. Hurwitz and W.G. Madow. _Sample Survey Methods and Theory, Volume II_. New York: John Wiley & Sons, 1953.

APPENDIX A

SAMPLING DISTRIBUTIONS: BIAS AND EFFICIENCY

A histogram or "picture" of the population appears in Figure 1.

Figure 1



The characteristics of interest are (a) the population mean (or average), $\overline{Y}$, where

$$\overline{Y} = \frac{2 + 6 + 8 + 10 + 12}{6} = 8$$

and (b) the population variance, σ^2, where

$$\sigma^2 = \Sigma_{i=1}^{N} \frac{(Y_i - \overline{Y})^2}{N} = \frac{64}{6} = 10.67$$

The task of the survey is, on the basis of a sample of a particular design and of a given size, to estimate the unknown value of the population mean $\overline{Y}$. To do this, a single sample of a given size is selected according to the desired design and an estimate constructed from the sample measurements obtained. Hopefully, the estimate will be "close" to the true value, though it will be impossible to determine, in any actual situation exactly how close the estimate in fact is.

A.2.1 Simple Random Sampling

With Replacement (SRSWR)

Consider the sampling design whereby a sample of size two is selected from the six index cards allowing a card to be selected twice; in other words, SRSWR of size two from a population of size six.

In total, there are 36 possible samples using this scheme. Although tedious, it is possible to identify each of these samples, determine the corresponding values for the selected units, and compute the sample mean for each sample as displayed in Table 1 together with the sample values, and the computed sample mean.

TABLE 1

SRSWR: n=2, N=6

Sample #	Sample Units	Sample Values	$\overline{y}$	Sample #	Sample Units	Sample Values	$\overline{y}$
1	AA	2, 2	2	19	DA	10, 2	6
2	AB	2, 6	4	20	DB	10, 6	8
3	AC	2, 8	5	21	DC	10, 8	9
4	AD	2, 10	6	22	DD	10, 10	10
5	AE	2, 10	6	23	DE	10, 10	10
6	AF	2, 12	7	24	DF	10, 12	11
7	BA	6, 2	4	25	EA	10, 2	6
8	BB	6, 6	6	26	EB	10, 6	8
9	BC	6, 8	7	27	EC	10, 8	9
10	BD	6, 10	8	28	ED	10, 10	10
11	BE	6, 10	8	29	EE	10, 10	10
12	BF	6, 12	9	30	EF	10, 12	11
13	CA	8, 2	5	31	FA	12, 2	7
14	CB	8, 6	7	32	FB	12, 6	9
15	CC	8, 8	8	33	FC	12, 8	10
16	CD	8, 10	9	34	FD	12, 10	11
17	CE	8, 10	9	35	FE	12, 10	11
18	CF	8, 12	10	36	FF	12, 12	12

From the set of sample means, a frequency table can be constructed.

Using the data from Table 3, we can compute characteristics of this sampling distribution, in particular the mean (or expected value) and the variance.

1. Mean

$$E(\bar{y}) = \frac{\Sigma \bar{Y}i\ fi}{36} = \frac{288}{36} = 8$$

Note that the mean, or expected value of the sample mean is 8, exactly to the mean of the original population, illustrating that the sample mean is an unbiased estimator of the population mean.

2. Variance

$$\mathrm{Var}(\bar{y}) = \frac{\Sigma \bar{Y}i^2 fi}{36} - \left\{ \frac{\Sigma Yi\ fi}{36} \right\}^2$$

$$= \frac{2496}{36} - \left(\frac{288}{36}\right)^2$$

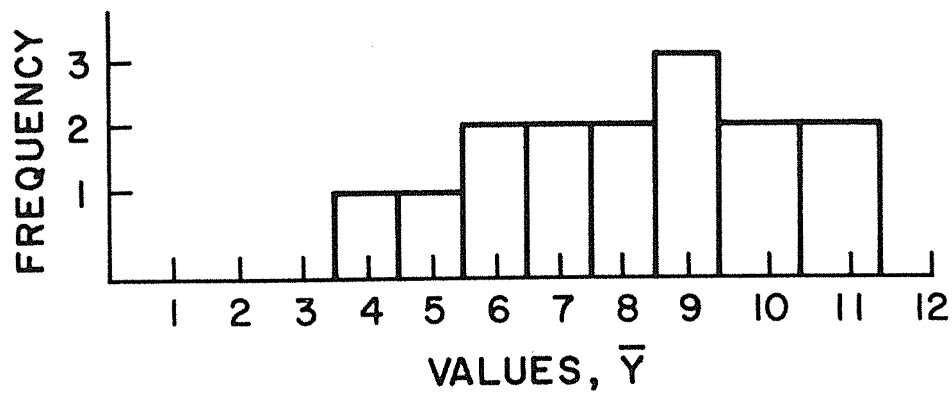
$$= 5.33$$

Recall that the variance of the original population was $\sigma^2 = 10.67$. Theoretically for SRSWR, the variance of the sampling distribution of the sample mean should be σ^2/n where n is the size of the sample selected; here the variance of the sampling distribution is 5.33 (10.67/2) which agrees with the theory. The crucial point is that the variation in the sampling distribution is reduced, relative to the original population, by a factor of two when sampling is carried out using this design.

Without Replacement (SRSWOR)

Consider now the effect of changing the design of our sampling procedure. Instead of allowing the repetition of sampled units, the sampling is restricted so that no card, once selected, can be selected a second time. For samples of size two from the population of size six, we thus have only 15 distinct samples, The corresponding histogram is shown in Figure 3.

Figure 3



As before, compute the mean (expected value), and the variance of the sampling distribution.

1. Mean

$$E\ (\overline{y}) = \frac{\Sigma \overline{Y}i\ fi}{15} = \frac{120}{15} = 8$$

Again note that the sample mean is unbiased for the population mean.

2. Variance

$$\text{Var}\ (\overline{y}) = \frac{1024}{15} - \frac{(120)}{15}^2 = 4.267$$

With this SRSWOR design, theory indicates that the variance of the sampling distribution is given by

$$V\ (\overline{y}) = \frac{\sigma^2}{n}\ \frac{(N-n)}{N-1}$$

TABLE 5

SRSWR: n=3, N=6

$\bar{y}i$	Frequency fi	$\bar{y}i$ fi	$\bar{y}i^2$fi
2	1	2	4
3 1/3	3	10	33
4	3	12	48
4 2/3	9	42	196
5 1/3	9	48	256
6	16	96	576
6 2/3	21	140	932
7 1/3	27	198	1,451
8	28	224	1,792
8 2/3	27	234	2,026
9 1/3	27	252	2,351
10	23	230	2,300
10 2/3	15	160	1,706
11 1/3	6	68	770
12	1	12	144
	216	1,728	14,585

1. Mean

The mean (or expected value) is thus found to be:

$$E\ (\bar{Y}) = \frac{1728}{216} = 8,$$

and hence, as before, we have shown that the sample mean is an unbiased estimator for the population mean.

2. Variance

The variance of the sampling distribution is computed as:

$$V\ (\bar{Y}) = \frac{14592}{216} - \left(\frac{1728}{216}\right)^2$$

$$= 3.55$$

which again agrees with the value of the variance suggested by theory, i.e., $\frac{\sigma}{n}$ or $\frac{10.67}{3} = 3.55$.

TABLE 6

Sample #	Sample Units (Green)	(Red)	Sample Values	$\overline{Y}$
1	A	D	2, 10	6
2	A	E	2, 10	6
3	A	F	2, 12	7
4	B	D	6, 10	8
5	B	E	6, 10	8
6	B	F	6, 12	9
7	C	D	8, 10	9
8	C	E	8, 10	9
9	C	F	8, 12	10

The sampling distribution of the mean is thus shown by the following frequency table.

TABLE 7

$\overline{Y}$	Frequency fi	$\overline{Y}$i fi	$\overline{Y}i^2fi$
6	2	12	72
7	1	7	49
8	2	16	128
9	3	27	243
10	1	10	10
	9	72	592

The expected value of the sample mean is again

$$E\ (\overline{y}) = \frac{72}{9} = 8$$

and the variance of the sample mean is

$$V\ (\overline{y}) = 1.78.$$

As before, the sample mean is unbiased for the population mean, and the variance has been substantially reduced through the simple device of stratification.

and so the sample variance is here, in SRSWR, unbiased for σ^2.

This illustrates the interesting fact that the way in which the sample is selected alters properties of the estimator, most importantly the degree of bias in estimating variances. This appendix does not provide a complete discussion of sampling theory, but does illustrate the main considerations involved in designing a sample, namely bias and efficiency.

APPENDIX B

ENERGY FOR HOMEOWNERS CASE: STUDY

Critical Path
Energy and the Homeowner

January 15th - February 15th	- Literature search and collection of relevant survey instruments - Questionnaire design
February 15th - March 1st	- Peer review of instrument (colleagues and potential users) - Sample design
March 1st - March 15th	- Revisions to questionnaire - Sample selection
March 15th - March 22nd	- Questionnaire pre-test (usually using a physical drop to a convenient sample, which need not be randomized for most mail surveys)
March 22nd - April 1st	- Evaluation of pre-test, revisions to questionnaire - Preparation of respondent file, assignment of unique identification numbers, and mailout envelopes
April 1st - April 15th	- Printing of questionnaires, covering letter and follow-up material (remember to print at least 1 1/2 times the number of questionnaires for the first mailout and second follow-up)
April 15th	- First mail-out
April 22nd	- Second follow-up (postcard) - Data entry of incoming responses
May 7th	- Third follow-up, with questionnaires to non-respondents ONLY - Data entry of incoming questionnaires
May 15th	- Fourth follow-up (postcard) (optional)
June 1st	- Close of survey
June 15th	- Completion of data entry and verification
June 15th - June 30th	- Analysis
July 31st	- Report due

ENERGY AND THE WINNIPEG HOMEOWNER

PRELIMINARY REPORT

by

G. Mason

D. Brown

July, 1982

B.3.1 Survey Type

Social scientists can select from among face-to-face interviews, telephone surveys and mailout surveys. Since the questionnaire did not require much intimate information and could be completed by more than one person over a period of time, the mailout survey is acceptable. In addition, it tends to be the most economical, an important consideration when public funds are used. Several potential disadvantages of mailout surveys include low response rates (15-25% is typical), and the inability for the field interviewer to confirm data. In this particular analysis, the advantages outweighed the risks and the mailout survey was selected.

B.3.2 Sample

Since the survey was designed to probe for homeowner's energy conservation activity, the sample was drawn from the 1981 City of Winnipeg property tax file by the City Planning Department. A systematic random sample of 503 single detached houses (no apartments, condominiums, or rowhouses) were selected. Of these, 357 eventually responded, constituting a 71.4 percent response rate. This is considerably higher than is normal in mailout surveys, but not unusual provided care is taken with the questionnaire format and an extensive follow-up is used. The follow-ups consisted of a postcard sent about a week after the initial package was mailed. After three weeks a reminder letter including a second copy of the survey was sent by special delivery, thereafter followed by a postcard one week later. Figure 1 shows the pattern of response over the sampling period. The effect of follow-up contact is clearly seen and accounts for about 40 percent of the total response.

B.3.4 Analysis

In the purpose of this paper, which is to review the broad results, the analysis need only represent the data in general terms. For this reason only tables and graphs are employed. Subsequent papers will include detailed statistical analyses of the factors which lead people to convert from oil to other forms of heat.

B.4 DESCRIPTION OF THE SAMPLE

The sample was drawn from the population of Winnipeg homeowners (as defined by the 1981 property tax roll for single detached homes). To the extent that some people had moved or otherwise were no longer at the home, the survey failed to make contact with a representative group. However, this accounts for about nine percent of the total non-response, or only 13 out of the total 146 non-responders.

Also, because only homeowners are surveyed, income and educational levels, as shown in Tables 1 and 2, tend to be higher than were the entire population of Winnipeg surveyed.

TABLE 3

Home Value of Respondents
(Percentage of Respondents in Each Category, Preliminary Figures)

0 - 19,999	0.9	$80,000 - 89,999	5.2
$20,000 - 29,999	1.7	$90,000 - 99,999	3.8
$30,000 - 39,999	9.0	$100,000 - 109,999	4.0
$40,000 - 49,999	10.4	$110,000 - 149,999	3.5
$50,000 - 59,999	15.9	$150,000 and over	0.9
$60,000 - 69,999	22.8	Missing	12.7
$70,000 - 79,999	9.2		

B.5 ATTITUDES TOWARD ENERGY ISSUES

Perhaps the most interesting aspect of the survey for both the general reader and policy planner are the attitudes toward current energy issues. These were addressed in the first part of the survey and the questions and response patterns appear below.

6. Both levels of government should share equally in this revenue from gasoline. N/A: 4.6%	12%	40%	12%	23%	9%
7. If a government approved insulating material is proved unsafe, then it is the responsibility of the government to pay all costs in making safe a home which is insulated with this product. N/A: 2.6%	45%	30%	8%	11%	2%
8. New dams to produce hydro-electricity should be paid for by increasing hydro rates by 15%, rather than by increasing the public debt. N/A: 2.6%	10%	31%	24%	25%	8%
9. If we have another energy crisis and fuel supplies became short, government should ration fuel rather than let the price rise. N/A: 4.0%	17%	45%	14%	16%	4%
10. As a measure to increase employment, government policy should allow industry to pay lower electricity rates than homeowners. N/A: 3.5%	4%	27%	11%	38%	16%

Most respondents (79%) agreed (or strongly agreed) that the energy industry should be owned by Canadians. Also most (85%) felt that since we have abundant supplies, our price should be lower than in areas where energy supplies are scarce.

Many respondents (73%) were in opposition to the large share of gasoline prices which are taken in taxes and this opposition was reflected in who should take this revenue.

TABLE 5

Expenditures on Energy Conservation

SECTION IV

Now, we would like to get some idea of how much money you have spent on conservation activities and how you financed them.

1. Approximately how much have you spent on energy conservation (do not include changes to the furnace), on your home since you bought it?

a.	$100 or less	13%	
b.	$ 101 - $ 250	6%	Approx. Mean = 1,117
c.	$ 251 - $ 500	15%	
d.	$ 501 - $1,000	16%	Standard Error = $52.07
e.	$1,001 - $1,500	15%	
f.	$1,501 - $2,000	10%	N/A: 3.2%
g.	$2,001 - $2,500	7%	
h.	over $2,500	14%	

Most respondents paid for this out of savings (51%) or small expenditures from monthly incomes (34%) and/or the Canadian Home Insulation Program (25%). In general, the larger the expenditure, the more likely was a personal loan from a bank or CHIP to be used to assist in the financing. Only 4 percent of the respondents indicated that they had made no expenditure at all on energy conservation. In most instances, these would be homeowners who had recently acquired a new home or a house which had been "retro-fitted."

B.7 ENERGY CONVERSION

Most homeowners had gas heat when they first purchased their home but 54 reported a conversion from oil to gas or electricity since they purchased their home. Table 6 shows the source of heat when the home was first purchased and the form of heating presently used.

ENERGY POLICY FOR HOMEOWNERS

FIELD OPERATIONS

July 1982

Two weeks followed with no further mailouts. The response pattern can be seen in attached literature.

Thus, it can be seen that by the time the third mailout was mailed on June 22, arriving at most houses on June 23, response had again dropped close to zero.

After much discussion about the most effective and efficient way of jolting non-responders into returning their surveys, the third mailout was sent special delivery. It contained a survey, a slightly more insistent covering letter (see attached) and a pre-stamped envelope. It was sent to approximately 230 homeowners who had so far not indicated they were unwilling to participate.

The third mailout was also quite successful. Response rose to 9 on June 24, 21 on June 25, 14 on June 28, and 17 on June 29. Several of the surveys which were returned after the third mailout were the copies sent in the first mailout, distinguishable by the absense of the "Printing Services" mark on the back of the survey. Older copies of the survey continued to be returned as the survey progressed.

The final mailout, a postcard similar to the first postcard but with a different date, was to be mailed on Tuesday, June 29, but was actually mailed on Monday, June 28. The reason for the discrepancy was a typographical error in the phone number supplied on the covering letter of the third mailout. The error may have prevented many people from phoning the Institute, and it was decided to send the postcard a day early to supply the correct phone number. It is doubtful whether the incorrect phone number affected the response rate, since the vast majority of those who contacted the Institute after receiving a survey were phoning to say that they would not do the survey.

the layout of the questions and did not answer correctly. For example, many indicated that there were two adults working, apparently assuming that this was the question, rather than giving occupations. A similar problem occurred with the education question. For example, a letter would be circled rather than a space checked, with no indication of how many adults, or which adults, had a certain level of education. People were also confused by the elementary/high school distinction, and next time the insertion of a "junior high school" designation would be useful. As it was, grade 8 and up were coded as "high school", and education to grade 7 coded as "elementary". Occupations were occasionally difficult to code also, as many answers were too general to fit into CCDO definitions. If an occupation was impossible to code, it received a "0007". For some occupations, the CCDO was modified slightly, as shown in the codebook. For example, if the occupation listed was "teacher", the number given was 2733, meaning "secondary school teacher", simply because the respondents' answers were not always as specific as were the CCDO listings.

The responses were coded and stored in computer files. Socio-economic indexes were added for each family.

SECTION I

Since 1974 the federal and provincial governments have been very active in energy policy. These first questions are designed to provide them with information on how you feel about parts of this energy policy. Please circle the number for each question that most closely reflects your views.

	STRONGLY AGREE 1	AGREE 2	UNDECIDED 3	DISAGREE 4	STRONGLY DISAGREE 5
1. The energy industry in Canada should be owned by Canadians.	1	2	3	4	5
2. Since Canada has abundant energy supplies, Canadians should pay a lower price for fuel than people in countries where energy is scarce.	1	2	3	4	5
3. The government's present policy of taking 90% of the price of every gallon of gasoline in taxes is justified.	1	2	3	4	5
4. The federal government should get most of this revenue from gasoline.	1	2	3	4	5
5. The provincial government should get most of this revenue from gasoline.	1	2	3	4	5
6. Both levels of government should share equally in this revenue from gasoline.	1	2	3	4	5
7. If a government approved insulating material is proved unsafe, then it is the responsibility of the government to pay all costs in making safe a home which is insulated with this product.	1	2	3	4	5
8. New dams to produce hydro-electricity should be paid for by increasing hydro rates by 15%, rather than by increasing the public debt.	1	2	3	4	5

	STRONGLY AGREE 1	AGREE 2	UNDECIDED 3	DISAGREE 4	STRONGLY DISAGREE 5
9. If we have another energy crisis and fuel supplies become short, government should ration fuel rather than let the price rise.	1	2	3	4	5
10. As a measure to increase employment, government policy should allow industry to pay lower electricity rates than homeowners.	1	2	3	4	5

SECTION II

Government energy policy has concentrated upon assisting homeowners to install insulation. The next set of questions relate to your present home, and its level of insulation when you bought it.

a. In what year did you buy this house? ____________

b. Which of the following energy conservation features did your house have at the time you first bought it?

1. Storm Doors
   - a. NONE
   - b. SOME
   - c. ALL
   - d. CAN'T RECALL
2. Storm Windows
   - a. NONE
   - b. SOME
   - c. ALL
   - d. CAN'T RECALL
3. Triple Pane Windows
   - a. NONE
   - b. SOME
   - c. ALL
   - d. CAN'T RECALL

3. Added triple pane windows
   a. NONE
   b. SOME
   c. ALL
4. Added or upgraded weather stripping on exterior doors
   a. NONE
   b. SOME
   c. ALL
5. Added or upgraded weather stripping on windows
   a. NONE
   b. SOME
   c. ALL
6. Added or upgraded attic insulation
   a. NONE OR MINIMAL (SAWDUST, HORSEHAIR, ETC.)
   b. 4" OF FIBERGLASS OR CELLULOSE
   c. 8" OF FIBERGLASS OR CELLULOSE
   d. 10" OF FIBERGLASS OR CELLULOSE
   e. MORE THAN 10" OF FIBERGLASS OR CELLULOSE
7. Added or upgraded wall insulation
   a. NONE OR MINIMAL (SAWDUST, HORSEHAIR, ETC.)
   b. 4" OF FIBERGLASS OR CELLULOSE
   c. 8" OF FIBERGLASS OR CELLULOSE
   d. 10" OF FIBERGLASS OR CELLULOSE
   e. MORE THAN 10" OF FIBERGLASS OR CELLULOSE
8. Added or upgraded basement insulation
   a. NONE OR MINIMAL (SAWDUST, HORSEHAIR, ETC.)
   b. 4" OF FIBERGLASS OR CELLULOSE
   c. 6" OF FIBERGLASS OR CELLULOSE
   d. 10" OF FIBERGLASS OR CELLULOSE
   e. NOT APPLICABLE (NO BASEMENT)

9. Please indicate what energy management devices have been installed since you bought this house. (Do not include added or upgraded insulation).
   a. HUMIDIFIER/DEHUMIDIFIER
   b. ADDED ATTIC VENTILATORS AND/OR FANS
   c. AUTOMATIC THERMOSTAT (TO CHANGE TEMPERATURES DURING THE DAY)
   d. DEFLECTORS ON AIR VENTS
   e. AWNINGS ON WINDOWS
   f. FRESH AIR INTAKE TO EXISTING FURNACE
   g. OTHER (SPECIFY)____________________

SECTION IV

Now, we would like to get some idea of how much money you have spent on conservation activities and how you financed them.

1. Approximately how much have you spent on energy conservation (Do not include changes to the furnace), on your home since you bought it?
   a. $100 OR LESS
   b. $101 - $250
   c. $251 - $500
   d. $501 - $1,000
   e. $1,001 - $1,500
   f. $1,501 - $2,000
   g. $2,001 - $2,500
   h. OVER $2,500

6. What is its approximate square footage to the nearest 100 sq. ft.?

____

7. What type of house is it?

a. BUNGALOW

b. 1 1/2 STOREY

c. 2 STOREY

d. 3 STOREY

e. BI-LEVEL

f. SPLIT LEVEL

g. OTHER (SPECIFY)__________________

8. Does it have a full basement? (Circle one)

a. YES

b. NO

9. What is the typical setting of the thermostat during the winter?

a. DAY (DEGREES FAHRENHEIT)______________

b. NIGHT (DEGREES FAHRENHEIT)______________

10. What type of water heater do you have? (Circle one)

a. GAS

b. ELECTRIC

c. OTHER (SPECIFY)______________________

11. To the nearest $5,000.00 what is the estimated value of this house?

____

SECTION VI

Finally, we would like to ask you some questions about your household. This will allow us to compare different households and precisely define to what extent present energy policies work.

1. What are the occupations of the adults (people 18 and over) who are full time residents in this house?

a. ADULT 1 ______________________

b. ADULT 2 ______________________

c. ADULT 3 ______________________

d. ADULT 4 ______________________

e. ADULT 5 ______________________

2. What is the highest education level attained by members of this household? (Check appropriate space).

	ADULT 1	ADULT 2	ADULT 3	ADULT 4	ADULT 5
a. ELEMENTARY	___	___	___	___	___
b. HIGH SCHOOL	___	___	___	___	___
c. SOME UNIVERSITY	___	___	___	___	___
d. SOME TECHNICAL SCHOOL	___	___	___	___	___
e. UNIVERSITY GRADUATE	___	___	___	___	___
f. TECHNICAL SCHOOL GRADUATE	___	___	___	___	___

3. Since children add greatly to energy costs please indicate the number of children you have in each age group normally living in your home. (If none, write "0").

________UNDER 5 YEARS OF AGE

________5 TO 13 YEARS OF AGE

________14 TO 18 YEARS OF AGE