

QUANTITATIVE DATA ANALYSIS AND THE RESEARCH PROCESS

In this section, the way in which quantitative data analysis fits into the research process – specifically the process of quantitative research – will be explored. As we will see, the area covered by this book does not solely address the question of how to deal with quantitative data, since it is also concerned with other aspects of the research process that impinge on data analysis.

Figure 1.1 provides an illustration of the chief steps in the process of quantitative research. Although there are grounds for doubting whether research always conforms to a neat linear sequence (Bryman 1988a, 1988b), the components depicted in Figure 1.1 provide a useful model. The following stages are delineated by the model.

Theory

The starting point for the process is a theoretical domain. Theories in the social sciences can vary between abstract general approaches (such as functionalism) and fairly low-level theories to explain specific phenomena (such as voting behaviour, delinquency, aggressiveness). By and large, the theories that are most likely to receive direct empirical attention are those which are at a fairly low level of generality. Merton (1967) referred to these as theories of the middle range, to denote theories that stood between general, abstract theories and empirical findings. Thus, Hirschi (1969), for example, formulated a 'control theory' of juvenile delinquency which proposes that delinquent acts are more likely to occur when the child's bonds to society are breached. This theory in large part derived from other theories and also from research findings relating to juvenile delinquency.

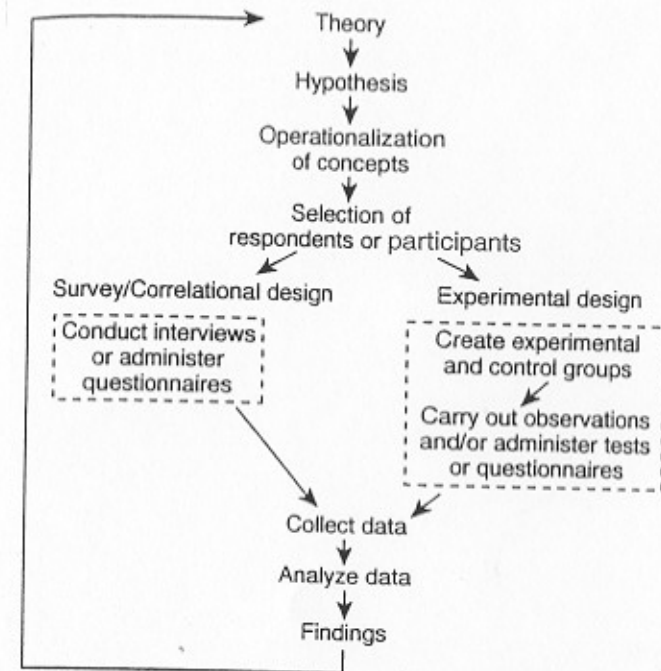


Figure 1.1 The research process

Hypothesis

Once a theory has been formulated, it is likely that researchers will want to test it. Does the theory hold water when faced with empirical evidence? However, it is rarely possible to test a theory as such. Instead, we are more likely to find that a hypothesis, which relates to a limited facet of the theory, will be deduced from the theory and submitted to a searching enquiry. For example, Hirschi, drawing upon his control theory, stipulates that children who are tied to conventional society (in the sense of adhering to conventional values and participating or aspiring to participate in conventional values) will be less likely to commit delinquent acts than those not so tied. Hypotheses very often take the form of relationships between two or more entities – in this case commitment to conventional society and juvenile delinquency. These 'entities' are usually referred to as 'concepts'; that is, categories in which are stored our ideas and observations about common elements in the world. The nature of concepts is discussed in greater detail in Chapter 4. Although hypotheses have the advantage that they force researchers to think systematically about what they want to study and to structure their research plans accordingly, they exhibit a potential

disadvantage in that they may divert a researcher's attention too far away from other interesting facets of the data he or she has amassed.

Operationalization of concepts

In order to assess the validity of a hypothesis it is necessary to develop measures of the constituent concepts. This process is often referred to as *operationalization*, following expositions of the measurement process in physics (Bridgman 1927). In effect, what is happening here is the translation of the concepts into variables; that is, attributes on which relevant objects (individuals, firms, nations, or whatever) differ. Hirschi operationalized the idea of commitment to conventional society in a number of ways. One route was through a question on a questionnaire asking the children to whom it was to be administered whether they liked school. Delinquency was measured in one of two ways, of which one was to ask about the number of delinquent acts to which children admitted (i.e. self-reported delinquent acts). In much experimental research in psychology, the measurement of concepts is achieved through the observation of people, rather than through the administration of questionnaires. For example, if the researcher is interested in aggression, a laboratory situation may be set up in which variations in aggressive behaviour are observed. Another way in which concepts may be operationalized is through the analysis of existing statistics, of which Durkheim's (1952/1898) classic analysis of suicide rates is an example. A number of issues to do with the process of devising measures of concepts and some of the properties that measures should possess are discussed in Chapter 4.

Selection of respondents or participants

If a survey investigation is being undertaken, the researcher must find relevant people to whom the research instrument that has been devised (e.g. self-administered questionnaire, interview schedule) should be administered. Hirschi, for example, randomly selected over 5,500 schoolchildren from an area in California. The fact of random selection is important here because it reflects a commitment to the production of findings that can be generalized beyond the confines of those who participate in a study. It is rarely possible to contact all units in a population, so that a *sample* invariably has to be selected. In order to be able to generalize to a wider population, a *representative sample*, such as one that can be achieved through random sampling, will be required. Moreover, many of the statistical techniques to be covered in this book are *inferential statistics*, which allow the researcher to demonstrate the probability that the results deriving from a sample are likely to be found in the population from which the sample was taken, but only if a random sample has been selected. These issues are examined in Chapter 6.

Setting up a research design

There are two basic types of research design that are employed by psychologists and sociologists. The former tend to use *experimental* designs in which the researcher actively manipulates aspects of a setting, either in the laboratory or in a field situation, and observes the effects of that manipulation on experimental participants. There must also be a 'control group' which acts as a point of comparison with the group of participants who receive the experimental manipulation. With a *survey/correlational* design, the researcher does not manipulate any of the variables of interest and data relating to all variables are collected simultaneously. The term *correlation* also refers to a technique for analyzing relationships between variables (see Chapter 8), but is used in the present context to denote a type of research design. The researcher does not always have a choice regarding which of the two designs can be adopted. For example, Hirschi could not *make* some children committed to school and others less committed and observe the effects on their propensity to commit delinquent acts. Some variables, like most of those studied by sociologists, are not capable of manipulation. However, there are areas of research in which topics and hypotheses are addressed with both types of research design (e.g. the study of the effects of participation at work on job satisfaction and performance – see Bryman 1986; Locke and Schweiger 1979). It should be noted that in most cases, therefore, the nature of the research design – whether experimental or survey/correlational – is known at the outset of the sequence signified by Figure 1.1, so that research design characteristics permeate and inform a number of stages of the research process. The nature of the research design has implications for the kinds of statistical manipulation that can be performed on the resulting data. The differences between the two designs are given greater attention in the next section.

Collect data

The researcher collects data at this stage, by interview, questionnaire, observation, or whatever. The technicalities of the issues pertinent to this stage are not usually associated with a book such as this. Readers should consult a textbook concerned with social and psychological research methods if they are unfamiliar with the relevant issues.

Analyze data

This stage connects very directly with the material covered in this book. At a minimum, the researcher is likely to want to describe his or her participants in terms of the variables deriving from the study. For example, the researcher might be interested in the proportion of children who claim to have committed no,

just one, or two or more delinquent acts. The various ways of analyzing and presenting the information relating to a single variable (sometimes called *univariate analysis*) are examined in Chapter 5. However, the analysis of a single variable is unlikely to suffice and the researcher will probably be interested in the connection between that variable and each of a number of other variables, i.e. *bivariate analysis*. The examination of connections among variables can take either of two forms. A researcher who has conducted an experiment may be interested in the extent to which experimental and control groups differ in some respect. For example, the researcher might be interested in examining whether watching violent films increases aggressiveness. The experimental group (which watches the violent films) and the control group (which does not) can then be compared to see how far they differ. The techniques for examining differences are explored in Chapter 7. The researcher may be interested in relationships between variables – are two variables connected with each other so that they tend to vary together? For example, Hirschi (1969: 121) presents a table which shows how liking school and self-reported delinquent acts are interconnected. He found that whereas only 9 per cent of children who say they like school have committed two or more delinquent acts, 49 per cent of those who say they dislike school have committed as many delinquent acts. The ways in which relationships among pairs of variables can be elucidated can be found in Chapter 8. Very often the researcher will be interested in exploring connections among more than two variables, i.e. *multivariate analysis*. Chapter 9 examines such analysis in the context of the exploration of differences, while Chapter 10 looks at the multivariate analysis of relationships among more than two variables. The distinction between studying differences and studying relationships is not always clear-cut. We might find that boys are more likely than girls to commit delinquent acts. This finding could be taken to mean that boys and girls differ in terms of propensity to engage in delinquent acts or that there is a relationship between gender and delinquency.

Findings

If the analysis of data suggests that a hypothesis is confirmed, this result can be fed back into the theory that prompted it. Future researchers can then concern themselves either with seeking to replicate the finding or with other ramifications of the theory. However, the refutation of a hypothesis can be just as important in that it may suggest that the theory is faulty, or at the very least in need of revision. Sometimes the hypothesis may be confirmed in some respects only. For example, a multivariate analysis may suggest that a relationship between two variables pertains to only some members of a sample, but not others (e.g. women but not men, or younger but not older people). Such a finding will require a reformulation of the theory. Not all findings will necessarily relate directly to a hypothesis. With a social survey, for example, the researcher may collect data on topics whose relevance only becomes evident at a later juncture.

As suggested above, the sequence depicted in Figure 1.1 constitutes a model of the research process, which may not always be reproduced in reality. None the less, it does serve to pinpoint the importance to the process of quantitative research of developing measures of concepts and the thorough analysis of subsequent data. One point that was not mentioned in the discussion is the *form* that the hypotheses and findings tend to assume. One of the main aims of much quantitative research in the social sciences is the demonstration of *causality* – that one variable has an impact upon another. The terms *independent variable* and *dependent variable* are often employed in this context. The former denotes a variable that has an impact upon the dependent variable. The latter, in other words, is deemed to be an effect of the independent variable. This causal imagery is widespread in the social sciences and a major role of multivariate analysis is the elucidation of such causal relationships (Bryman 1988a). The ease with which a researcher can establish cause and effect relationships is strongly affected by the nature of the research design and it is to this topic that we shall now turn.