

8.2 Introduction

Before decisions can be made about how to collect data to answer research questions, consideration needs to be given to the kind of data required, where they will come from, how they will be selected, and how they will be analysed. Hence, this chapter discusses:

- *types* of data used by social researchers – primary, secondary and tertiary;
- *forms* in which data are produced in the social sciences – in either words or numbers, or both;
- *sources* of data in terms of the settings from which they will be obtained – natural social settings, semi-natural settings, artificial settings, and social artefacts;
- *selection* of data, with particular reference to sampling; and
- the role of *case studies* in social research.

8.3 Types of Data

Data used in social research can be of three main types: *primary*, *secondary* and *tertiary*. *Primary* data are generated by a researcher or researchers who is/are responsible for the design of the study from which the data come. These are 'new' data, used to answer specific research questions, and the researcher can describe why and how they were collected. *Secondary* data are raw data that have already been collected by someone else, either for some general information purpose, such as a government census or other official statistics, or from a specific research project. In both types of secondary sources, the original purpose in collecting such data could be different from that of the subsequent user, particularly in the case of a prior research project. *Tertiary* data have been analysed either by the researcher(s) who generated them or by a user of secondary data. In this case, the raw data may not be available, only the results of this analysis (see Figure 7.1).

Types of Data

Primary

Generated by the researcher

Secondary

Generated by another researcher

Tertiary

Analysed by another researcher

While primary data can come from many sources, they are characterized by the fact that they are the result of direct contact between a researcher and the source. As researchers have control over the production and analysis of primary data, they are in a position to judge their quality.

Secondary data can come from the same kind of sources as primary data. The use of secondary data is often referred to as secondary analysis.

It is now considered best practice for sets of such data to be archived and made available for analysis by other researchers; they can be interrogated with

different research questions, to validate prior findings, or to use for purposes of comparison.

While there are obvious advantages in using secondary data, such as savings in time and costs, there are also limitations and disadvantages.

- The research is likely to have addressed a different research problem and questions, in a different social context.
- Prior studies may have used different ontological and epistemological assumptions, which may not be evident.
- Decisions may have been made in the coding that are inconvenient or problematic.
- It may be difficult to judge the quality of the data.
- The data may be dated (although this may not be a problem in historical, comparative or theoretical studies).

With tertiary data, a researcher is two steps removed from the original primary data. Published reports of research, and officially collected 'statistics', invariably include tables of data that have summarized, categorized or otherwise manipulated raw data. This is true of most government census reports, where access to the original dataset may not be possible. When government agencies or other bodies do their own analysis on a census, they produce genuine tertiary data.

8.4 Forms of Data

Data end up in two main forms, as *numbers* or *words*. Even if they start out as sound or video recordings, for example, when prepared for analysis they will be either in words or numbers, or perhaps both.

While there is a common prejudice in favour of numerical data as being necessary for 'objective', 'scientific' research, data in non-numerical forms are now also generally accepted. As we shall see later in this chapter, the distinction between words and numbers, between qualitative and quantitative data, is not a simple one (see Figure 7.1).

Primary data usually start out as words; for example, as written questions and answers, including pre-coded categories, or recorded conversations. Some data are recorded in words, they remain in words throughout the analysis, and the findings are reported in words. The original words will be transformed and manipulated into other words, and these processes may be repeated more than once. While the level of the language may change, say from lay language to technical language, throughout the research the medium is always words. In

Forms of Data

Words

At source
During analysis
For reporting

Numbers

Soon after source
For analysis
For reporting

addition, some data may start out as images, such as photographs or video recordings. However, words must be used to describe and interpret such images before analysis can begin.

In other research, the initial communication, say on a questionnaire or between interviewer and research participant, will be transformed into numbers immediately, or prior to the analysis. The former involves the use of pre-coded response categories, and the latter the post-coding of answers provided in words, as in the case of open-ended questions. When numbers are attached to both sets of categories the subsequent analysis will be numerical. The findings of the research will be presented in numerical summaries and tables. However, words will have to be used to interpret and elaborate the numerical findings. Hence, in quantitative studies, data normally begin in words, and are then transformed into numbers, subjected to various kinds of statistical manipulation, and are reported in both numbers and words; from words to numbers and back to words.

The interesting questions here are: Whose words are used in the first place? And what process is used to generate them? In the case where answers are communicated using a predetermined set of categories, both questions and response categories will be in a researcher's words; a respondent *only* has to interpret both. However, this is a big 'only'. As Foddy (1993) and Pawson (1995, 1996) have pointed out, this is a complex process that requires much more attention and understanding than it has normally been given.

Sophisticated numerical transformations can occur at the data reduction stage of analysis. For example, responses to a set of attitude statements, in categories ranging from 'strongly agree' to 'strongly disagree', can be numbered, say, from 1 to 5. Subject to some test of unidimensionality, scores from each statement can be combined to produce a total score for a composite attitudinal scale. Such total scores are well removed from a respondent's original reading of the words in the statements and the recording of a response in a category with a label in words.

In studies that deal with large quantities of textual data, as in the case of transcribed in-depth interviews, manipulation can occur in two main ways: in the generation of categories in which segments of text are coded (as in grounded theory); or in the abstraction of technical language from lay language (as in the case of Schütz's and Giddens's use of Abductive logic). In both of these forms of analysis, it is possible to do simple counting, for example, to establish the number of times a category occurs in a body of text, or the number of respondents who can be located in each of a set of ideal types.

So far, this discussion of the use of words and numbers has been confined to the collection of primary data. However, such manipulations may have already occurred in secondary data, and will certainly have occurred in tertiary data.

The controversial issue here is the effect that any form of manipulation may have on the original data. To reiterate earlier arguments, we need to be clear at the outset that researchers cannot produce 'pure' data; data are constructed. If all observation is interpretation, and concepts are theoretically saturated, then some form of construction is involved from the very beginning. For example, even if a conversation is recorded unobtrusively, any attempt to understand what went on requires interpretation and the use of concepts. How much

manipulation occurs is a matter of choice; how conscious and intentional such manipulations are, and the degree to which they are documented and disclosed, are also salient considerations. Researchers who prefer to work with qualitative data through all stages of a study may argue that it is bad enough abstracting lay language into technical language without translating either of them into the language of mathematics. A common fear about such translations is that they end up abstracting or distorting the social world out of all recognition, with the result that research reports based on them become separated from the social reality on which they are based, becoming either meaningless or, possibly, dangerous if acted on. We shall encounter arguments in favour of quantification later in the chapter.

8.5 Sources of Data

Regardless of whether data are primary, secondary or tertiary, they can come from four different types of settings and interactions within those settings: *natural social settings*, *semi-natural settings*, *artificial settings* and from *social artefacts* (see Figure 7.1). In research conducted in a *natural social setting*, researchers enter an area of social activity and study people going about their everyday lives. In a *semi-natural setting*, individuals are asked to report on their activities as they occurred in natural settings, while in an *artificial setting* social activity is contrived for experimental or learning purposes. The fourth kind of social setting involves the examination of records or traces, or other 'available data', left by individuals or groups. A fundamental distinction here is whether social activity is studied as it occurs *in situ*, or whether it will be artificially or historically reconstructed in some way. This fourth type of research setting introduces a time dimension, and this will be examined shortly.

Natural Social Settings

Natural social settings involve three main levels of analysis: micro-social, meso-social and macro-social.¹ These levels vary in scale from individuals and small social groups, through organizations and communities, to institutions and large-scale social situations, such as cities and regions, nations and multi/transnational bodies.

Micro-social phenomena

At its most basic level, the *micro-social* level is made up of individuals in their everyday social settings. However, unlike a great deal of research in psychology, this interest in the individual is often as a participant in some small-scale social unit rather than as an individual *per se*. Micro-social relations are normally characterized by face-to-face social interaction in which social actors give meaning to their own actions, and the actions of the others involved, and take these others into account when making decisions about their own actions. Many

Sources of Data

Natural social settings

Micro

- Individuals
- Small groups
- Social episodes

Meso

- Organizations
- Communities
- Crowds
- Social movements

Macro

- Social institutions
- Social structures
- Nations
- Multinational bodies

Quasi-experiments

Semi-natural settings

- Individuals' characteristics
- Individuals as informants
- Representative individuals
- Life histories
- Individuals as case studies

Artificial settings

- Experiments
- Simulation and games

Social artefacts

- Official statistics
- Public documents
- Private documents
- Personal records

or illegal. The social relationships in organizations are largely secondary in nature, with membership that may be either compulsory (e.g. in a prison) or voluntary (e.g. in a sporting club), full-time or part-time, paid or unpaid, long-term or short-term. While organizations change, at times rapidly, the structure of relations and forms of authority and leadership are likely to be relatively enduring.

'Community' has been included to refer to a diverse range of social phenomena. Like 'society', this concept has many uses. For example, we refer to 'local' communities, 'the' community, 'ethnic' communities, 'regional' communities and 'the world' community, all of which are very different. The notion of community being used here refers to looser forms of social organization in which either space or common interests are the defining characteristics. These social collectivities are different from organizations, and remove us one step further from primary and

such situations have a history and, usually, a relatively permanent membership; they develop and reproduce patterns, structures and institutions.

The other major type of micro-social phenomenon is the social episode; those social interactions that are limited in time and space, such as social gatherings of various kinds. Most of the participants in these situations may not meet again, or certainly not in the same circumstances or with the same membership. Intersubjective meanings develop and then dissipate, and the structuring of social relations is fleeting. The number of people involved in social episodes can vary considerably, from two to hundreds, thus making their inclusion under this category of social phenomenon somewhat arbitrary. However, the processes involved in understanding such phenomena are likely to be social-psychological in nature and similar to those used for small groups.

Meso-social phenomena

Meso-social phenomena include organizations, communities and crowds. As relatively permanent and large social groups with established goals, organizations can be public or private, for business or pleasure, legal

even secondary relationships. Forms of collective behaviour, such as crowds and social movements, can also be included here.

Macro-social phenomena

Macro-social phenomena are much larger social entities than those already discussed and are serious abstractions of some kind. Examples of such abstract phenomena are social institutions and social structures, both of which are the products of theorizing by experts. While institutions and structures are usually discussed in the context of a society, or nation,² it is possible to use them in the analysis of cross/inter/multinational bodies, such as transnational corporations and international non-government organizations.

At the macro level, it is becoming increasingly necessary in social science to move beyond nations, or even formal comparisons between nations, to deal with social phenomena that transcend borders. We are witnessing a developing paradox in which national and regional boundaries are becoming both less important and more important at the same time. In terms of world economics and migration, national boundaries have become less significant, but in terms of politics and social identity, national and regional boundaries are assuming greater significance.

It is difficult to use 'naturalistic' methods to study large-scale social units. In these cases, it may be possible to regard the unit of analysis as societal – as involving continuing social activity – but accept that it may not be possible to study this activity fully or directly.

The boundaries between micro, meso, and macro-social settings are not rigid; the three categories are intended to indicate a range of possible research sites that vary in terms of the number of people involved, the complexity of their social relationships, and their relative geographical concentration or dispersion and level of abstraction.

Quasi-experiments

The category of quasi-experiment is included under *natural social settings* to identify research in which experimental procedures are used outside the laboratory. Rather than contriving an experiment, researchers may use opportunities as they arise in natural settings. An example comes from a before-after study of a town in which a change of circumstances was introduced that was outside the researcher's control. Many years ago, Blaikie participated in a study of an isolated town on the west coast of the South Island of New Zealand. A thriving gold-mining centre in the late nineteenth century, the town was located on a narrow coastal strip between mountains and the sea. It was isolated at the end of a road. A decision was made to extend the road across a mountain pass to a popular tourist resort area, thus creating the potential for the town to be included in a tourist route. A project was designed to study the town before and sometime after the road was constructed to assess the impact of tourism. While the design of this study did not satisfy experimental ideals, the natural setting and fortuitous timing made a quasi-experiment possible. Many social impact studies are of this kind.

Semi-Natural Settings

Probably the most common form of research in the social sciences involves asking individuals to report on their own or other people's activities, attitudes and motives; or on social processes and institutionalized practices.

Individuals' characteristics

Three main kinds of data are collected in these studies: demographic characteristics; knowledge, attitudes and worldviews; and reported behaviour. Demographic characteristics are an essential part of data collected in censuses and social surveys. These include age, gender, marital status, education, income, occupation, place of residence, ethnic background and religion. While these characteristics may have social origins and social consequences, the analyses undertaken on them search for connections between such variables and other variables. The social processes or structures on which these connections might depend are left very much in the background, and assumptions about them may be made. For example, a respondent may be asked about their occupation, which is then used by the researcher to establish a status hierarchy. In everyday life, occupation might enter into the structuring of social relationships quite differently in different settings. What goes on in the natural setting may be reported selectively, ignored or distorted by the researcher.

The second common focus in studies of individuals is their perceptions, knowledge, attitudes, beliefs, values and such like. The aim is usually to use these data to explain reported behaviour. However, not only may such data be taken out of the context in which they might be relevant, but the connection between attitudes and behaviour may be extremely tenuous. Another major difficulty in asking people about themselves is the gap between what they say and what they actually think and do.

The third feature of this type of research is to ask individuals to report behaviour. This is either the individual's own behaviour (e.g. the frequency and duration of time spent on social media in, say, a week), or the behaviour of others (e.g. parents' assessment of the number of hours their children watched television in a particular week).

Individuals as informants or representatives

Individuals can also be studied as special persons (e.g. in the case of political biographies), as representatives of a particular type of social actor (e.g. 'indulgent parent' or 'rebellious teenager'), or as members of a particular social category (e.g. youth, pensioners, the working class, tertiary-educated).

Individuals as case studies

The study of single persons, perhaps as in-depth case studies, lies on the boundary of *social science*, particularly if the person's social context is given little or no attention. It is possible to assess an individual's perceptions of the social world

and to get them to report their social experiences; that is, their interaction with other people. When primary emphasis is placed more on cognitive processes, as in psychoanalytic studies of political leaders, the research may be more correctly classified as psychological or behavioural rather than social.

Artificial Settings

It is possible to place people in experimental or simulated conditions in order to study some form of social behaviour in a controlled environment.

Experiments

For many social scientists, to be able to hold some variables constant while others are manipulated, and then to observe the outcome, is considered to be the only way to explain any social phenomenon conclusively.³ All other research designs are regarded as deviations. In practice, the use of genuine experiments in the social sciences is limited to some specific fields, such as small groups; quasi-experimental studies are conducted in fields such as education and media studies. Hence, experiments are given only a brief treatment here. Rather, pseudo-experimental language is commonly used in other types of research, such that 'independent' and 'dependent' variables have become almost universally adopted in a great deal of social research.

It is in psychological and medical research that experiments are most commonly conducted.

The purpose of the simple experiment is to test whether a treatment causes an effect . . . To determine that a treatment causes an effect, you must create a situation where you show that the treatment comes before the effect, that the effect occurs after the treatment is introduced, and that nothing but the treatment is responsible for the effect . . . The ideal for doing a psychological experiment to demonstrate causality would be to find two identical groups of subjects, treat them identically except that only one group gets the treatment, test them under identical conditions, and then compare the behavior of the two groups. (Mitchell and Jolley 1992: 169)

Simple experiments are particularly suitable where a single cause is assumed to produce an effect. However, it is possible to extend an experimental design to deal with different levels of treatment, and with more than one kind of treatment.

There are a number of possible threats to the validity of the relationship between treatment and effect in experiments. This is commonly referred to as *internal validity*. See, for example, Campbell and Stanley (1963a: 5), Mitchell and Jolley (1992: 242–3) and Neuman (2014: 300–05) for reviews.

The effects of the experimental process on the subjects can also threaten the representativeness (*external validity*) of the results. However, one of the most serious threats to the possibility of generalizing results obtained in social experiments comes from the fact that people may behave differently in experimental

situations than they do in natural situations. Also, participants in experimental settings may not be representative of broader populations.

Experimental research with human subjects also requires a very rigorous consideration of the ethical implications of the research design and procedures.

Simulation and games

Social life has been simulated for a number of reasons, perhaps the most common being for educational purposes. Such *simulations* or *games* allow the participants to:

- experience features of social life under controlled conditions;
- experience and be involved in initiating co-operation and conflict;
- experience what it is like to be a particular type of person; for example, to be wealthy or poor, to have high status or low status; and
- learn about how power is acquired and used.

However, it is the use of simulation and games as a way of modelling some aspect of social life, and as a research technique, that is of interest here.

Such games require a set of rules. While some factors are held constant by the controller of the activity, the participants are able to manipulate others in the course of their activity together. The effects of changing what is controlled and what can be manipulated can be observed and analysed. Therefore, games are a form of experimentation in artificial settings. However, they differ from experiments in that they attempt to replicate some real social situation; they may be less concerned with establishing causation and more concerned with recognizing and understanding the complexities of social processes.

The use of simulation as a method for investigating social phenomena has been facilitated in recent years by the advent of small powerful computers and developments in computer science and information technology. Now people need not be involved; social processes that are not directly accessible, or are of too large a scale to be observed directly, can be modelled. The logical possibilities of a set of specifications or assumptions can be explored when the values of certain variables are changed (see Gullahorn and Gullahorn 1963; Inbar and Stoll 1972; Hanneman 1988; Ragin and Becker 1989; Brent and Anderson 1990; Garson 1990; Anderson 1992; Gilbert and Doran 1993; Lee 1995; Gilbert 1995).

Social Artefacts

The fourth main source of data, *social artefacts*, involves neither natural nor artificial situations. Rather, it involves the traces of social activities that people leave behind. Records of past social activities can be found in various places. Some are kept officially – such as censuses, publicly available minutes of meetings, or biographies and autobiographies – while others have been kept for private purposes – such as the internal reports and correspondence of a company or organization, and diaries, private letters or family photographs and genealogies. Unlike public records, these

latter are defined as ‘private’ because there is no legal obligation to provide open access to them. In addition, some individuals keep personal records, such as diaries or journals, which, during their lifetime, are intended for no one else’s eyes but the author’s. It is usually only after their death, or by special permission, that access to them can be gained. Other social artefacts can be found in such places as cemeteries, maps, land title records, photographs, sound recordings and moving images.

While social artefacts are the sources on which historians have to rely, they can also be invaluable, and in some cases are the only data available, to social scientists.

A Dilemma

In designing social research, researchers need to choose whether to use their own theoretically informed concepts and views of the world, or whether to take as their starting-point how participants and informants view and understand their world. There are essentially two choices here: one is for a researcher to define the phenomenon in terms of the technical concepts and ideas of some theoretical perspective within a discipline; and the other is to work with social actors’ construction of reality, at least as the starting-point for the investigation. For example, what constitutes a social episode or a social group is not simply a matter of observation; it is either a product of the way social actors consider their social world to be organized, or the result of a social scientist’s way of viewing the social world; it is either a *social* construction or a *sociological* construction. In both cases, information may be sought from the social actors; the difference is how that information is regarded.

Abstract concepts, such as social institutions and social structures, are usually *sociological* constructions, the inventions of social scientists. Nevertheless, as with social groups and social processes, data can be obtained both from researchers’ observations and experience, as well as from social participants’ knowledge, perceptions and experiences. Social actors will invariably see their world differently than do social scientists. For example, social scientists have conceptualized structures of inequality in various ways, such as three social classes (upper, middle and lower) or as a hierarchy of occupations, but the social actors concerned may not share these conceptions. They may have a very different way of ‘structuring’ their world (e.g. in terms of ‘insiders’ and ‘outsiders’, or as a set of concentric circles around their own position in society). Differences between *social* constructions and *sociological* constructions pose a fundamental methodological problem for researchers. The question is whose construction of reality should provide the foundation for understanding social life. It is on this issue that the four logics of inquiry adopt different positions.

8.6 Selection of Data

All social research involves decisions about how to select data from whatever the source or sources may be. This is true regardless of the purposes of the research, the research setting, the time dimension, the type of social phenomenon

being studied, the type of data, the form of the data, and the methods of data collection, reduction and analysis. When data are obtained separately from a number of individuals, social units or social artefacts, the researcher has a choice of either attempting to study an entire target population – all of those that meet certain criteria – or selecting a sample from that population. When an entire population is used, it has to be selected by defining the criteria for inclusion – such as time, place and individual characteristics. When a sample is used, further selections are made from the defined population, using one or more sampling methods.⁴

When sampling is used, it is frequently the weakest and least understood part of a research design. The type of sample selected, and the method used to do so, can have a bearing on many other parts of a research design, and these decisions can determine the kind of conclusions that can be drawn from a study.

This section begins with a discussion of some of the key concepts in sampling, in particular the technical meanings of *population* and *sample*. This is followed by a review of the major sampling methods, commonly referred to as random and non-random, but identified here as probabilistic and non-probabilistic. The idea of ‘random’ methods has the disadvantage of being associated with ‘accidental’ sampling rather than the need for representativeness.

As part of this review, consideration is given to how different methods can be combined, and how the size of a sample can be established. In addition, some comments are made on the connection between the use of tests of significance and probability sampling methods. Details of the techniques used in the major methods of probability sampling, and the associated mathematics, will not be elaborated here. (See, for example, Kish 1995 and Scheaffer et al. 2006 for technical details.) Rather, an overview of the major sampling decisions is provided, and consideration is given to the implications that the choice of sampling method can have for other research design decisions.

Populations and Samples

First, it is necessary to clarify the technical meanings of *population* and *sample*. In order to apply a sampling technique, it is first necessary to define the target population from which the sample is to be drawn. A *population is an aggregate of all cases that conform to some designated set of criteria*. Population elements are single members or units of a population; they can be such things as people, social actions, social situations, events, places, times or documents. The researcher is free to define a population in whatever way is considered appropriate to address the research question(s). For example, a population might be defined as:

- the citizens of a country at a particular time;
- first-year university students at a particular university at a specific time;
- landline telephone subscribers in a particular city;
- people of a particular age in a particular setting or context;
- all the issues of a particular newspaper published in a specified twelve-month period;

- only the Saturday issues of this newspaper during this period; or
- only articles in these newspapers that report domestic violence.

A *census collects data from all population elements* and is used to describe the characteristics of the population.

A *sample is a selection of elements (members or units) from a target population and may be used to make statements about that population*. The ideal sample is one that provides a perfect representation of a population, with all the relevant features of the population included in the sample in the same proportions. However, this ideal is difficult to achieve. In a *probability sample, every population element must have a known and non-zero chance of being selected*. Most types of probability samples will also give every element an equal chance of being selected. Non-probability samples do not give every population element a chance of selection. The relationship between the size of the sample and the size of the population is the *sampling ratio*.

While sampling can introduce many complexities into the analysis of data, it is used for a variety of reasons. Studying a whole population may be slow and tedious; it can be expensive and is sometimes impossible; and it may also be unnecessary. Given limited resources, sampling can not only reduce the costs of a study, but, given a fixed budget, can also increase the breadth of coverage. Stratified sampling also has the advantage of comprehensively representing specific diversity in a target population.

Methods of Sampling

A range of methods for drawing a sample is available. While sampling methods aim to represent the population from which the sample is drawn, some are compromises on this ideal. The nature of the research, the availability of data, and cost will determine the choice of method. Sampling methods have been divided along two dimensions: probability versus non-probability, and single-stage versus multi-stage.

Single-stage probability sampling

Simple random sampling

If a decision is made to use probability sampling, then a choice of methods must be made.⁵ *Simple random sampling* is the standard against which all other methods are judged. It involves a selection process that *gives every possible sample of a particular size an equal chance of being selected*. However, even simple random sampling does not guarantee an exact representation of a population; it is possible to randomly draw very ‘biased’ or skewed samples. What simple random sampling does is allow the use of probability theory to provide an estimate of the likelihood of such samples being drawn. The other probability sampling methods provide different kinds of compromises on simple random sampling, each with its own advantages in terms of cost or convenience and degree of sacrifice in terms of accuracy.

Sampling Methods

Single-stage probability

- Simple random
- Systematic
- Stratified
- Cluster

Multi-stage

Single-stage non-probability

- Accidental/Convenience
- Quota
- Judgemental/Purposive
- Snowball
- Theoretical

Simple random samples require that each element of a population be identified and, usually, numbered. Once the size of the sample has been determined, a list of computer-generated columns of random numbers can be used to make the selection from the population.⁶ For samples between 10 and 99, two columns of numbers will be used; between 100 and 999, three columns; between 1,000 and 9,999, four columns, and so on. Not all combinations of digits in the selected columns will be relevant as they may lie outside the range of the desired sample size. However, by scanning the column, the numbers

that *are* relevant can be noted until the desired sample size is reached. These numbers then determine the population elements to be selected. This process can now be automated.

Systematic sampling

Systematic sampling provides a method that avoids having to number the whole population. If the population elements can be put in a list, they can be counted and a sampling ratio decided to produce the desired sample size; for example, one in five, or one in sixteen. In effect, the list is divided into equal zones the size of the denominator of the sampling ratio (e.g. five or sixteen), and one selection is made in each zone. The only strictly random aspect to this method of selection is determining which element in the first zone will be selected. Then it is a matter of counting down the list in intervals the size of the denominator (e.g. five or sixteen).

The systematic sampling procedure is simple and foolproof. However, it does have some potential dangers. Should the size of the zone correspond to a regular pattern in the list, the method may produce a very biased sample. For example, if houses in a 'grid iron' pattern of city blocks are being selected, with a sampling ratio of one in sixteen, and if there happened to be sixteen houses in each block, and if the first house was randomly selected in the first zone, then the sample would consist only of corner houses. If the subject of the study is traffic noise pollution, these corner houses are likely to be affected differently to others in the block. There are two ways to protect a sample from such bias. One is to make a random selection more than once, thus changing the selection within the zones from time to time. Another is to double or treble the size of the zones and then to make two or three random selections within each zone. It is desirable to use either or both of these methods as they introduce a greater degree of randomness into the selection.

Stratified sampling

Stratified sampling requires a population to be divided into one or more sets of categories and then sample selections are made within each category. This method is used for two main purposes, the first of which is to ensure that particular categories in the population are represented in the sample in the same proportion as in the population, for example, gender, age or ethnic identification. In order to do this, it must be possible to identify the relevant characteristic in the population. If this is possible, the population elements can be grouped into the desired categories, or strata, before selections are made. By using the same sampling ratio in each category or stratum – male and female or age groups – the population distribution on such characteristics will be represented proportionately in the sample. A second purpose is to ensure that there are sufficient numbers in the sample from all categories that are to be examined. For example, if people are to be compared in terms of their religious affiliation, and a particular minority religion is considered to be important for this purpose, simple random, systematic or even stratified sampling may produce insufficient numbers from this category for later analysis. Assuming that this is the only under-represented stratum, one solution is to lower the sampling ratio for this stratum; for example, 1:3 rather than, say, 1:15 for the other strata. Another solution is to draw equal numbers from each stratum, by varying the sampling ratios. Once the number that is required from each stratum is decided, the denominator of the sampling ratio is arrived at by dividing that number by the number in the stratum. For example, if there are four strata of 5,000, 1,500, 1,000 and 100, and 50 are required from each stratum, the sampling ratios would be 1:100, 1:30, 1:20 and 1:2.

It is important to note that if the same sampling ratios are *not* used in all strata, the resulting sub-samples cannot be combined for analysis, as every member of the population did not have an equal chance of being selected. The strata are, in effect, separate populations. This problem can be overcome by weighting each stratum in the sample to restore it to the original relative proportions of the population strata. In the example just used, as the ratios between the population strata are 50:15:10:1, the numbers on which the data in each stratum are based would need to be multiplied by its respective ratio figure, or some multiple of it. The main disadvantage with this procedure is that any lack of representativeness in the initial sample will be magnified in the 'reconstituted' sample, particularly in cases where large weights are used. This kind of weighting procedure can also be used in simple random sampling when the population proportions on some variable are known, and the sampling proportions differ considerably from these.

Cluster sampling

All other forms of probability sampling use one or a combination of simple random sampling, systematic sampling or stratified sampling. The fourth common sampling method is cluster sampling, one version of which is known as *area sampling*. A cluster is a unit that contains a collection of population elements. Cluster sampling selects more than one population element at a time, for

example, a classroom of students, a city suburb of households, a street or block of residences, a year of issues of a newspaper, or a month of applications for citizenship. Cluster sampling is generally used when it is impossible or very difficult to list all population elements. It also has the advantage that it can reduce the cost of data collection by concentrating this activity in a number of areas, rather than being scattered over a wide area. However, as clusters are unlikely to be identical in their distribution of population characteristics, cluster sampling will normally be less accurate than simple random sampling. It is not difficult to select a very biased set of clusters, and this problem is exacerbated if only a few are selected.

Multi-stage sampling

Cluster sampling is often the first stage, or perhaps one of the stages, of multi-stage sampling designs. The selection of the clusters themselves, and later selections within clusters, can use any of the three probability sampling methods. It is preferable that clusters are of equal size; otherwise each population element will not have an equal chance of being selected. However, this is not always achievable as natural clusters may vary considerably in size. One rather complex method for overcoming this is to stratify the clusters roughly according to size, and use a sampling ratio in each cluster that will give each population element a more or less equal chance of selection. Weighting can also be used, although this adds to the complexity.

Multi-stage sampling is commonly used in surveys of householders. For example, if householders in a large metropolitan area have been defined as the population, the first stage could be a random selection of administrative areas (e.g. local city councils), perhaps stratified by size, and/or mean socio-economic status of the residents. The second stage could be based on subdivisions that are used in census collection (perhaps using simple random procedures, or stratified procedures if the areas vary considerably in size). This could be followed by a random sample of households (perhaps using the systematic method while walking along each street). Finally, a member of the household could be selected by a random procedure. This sampling design does not require the identification of households and householders until the very last stages, and even then they need not be listed. Efficiencies in data gathering can be achieved as interviewers can concentrate their efforts and save a great deal of time and money in travelling. However, at each stage in this sampling design, sampling errors can creep in. In order to compensate for this, a larger than the minimum desirable sample could be used. Nevertheless, this design represents an example of how practical problems (e.g. not having access to lists of householders) can be overcome, and how a compromise can be struck between precision and cost. This is the kind of sample that was used in the first three of the sample research designs in chapter 12.

It is possible to combine probability and non-probability methods in a multi-stage sample. For example, the selection of initial clusters, such as areas of a city based on the demographic characteristics of the residents, could be based on a judgement of how each area typifies the range of these characteristics. Subsequent sampling in each cluster could then use probability methods. Judgemental sampling will be discussed shortly.

Single-stage non-probability sampling

So far we have concentrated on issues related to probability sampling. Such sampling may be necessary in order to answer certain kinds of research questions with large populations. However, in addition to the use of populations rather than samples, some studies either do not need to generalize to a population, or cannot adequately identify the members of a population in order to draw a sample. For example, research on people who are infected with HIV/AIDS faces the problem that lists (sampling frames) of such people are difficult to obtain or are not available. To insist on the use of random sampling would make the research impossible. Therefore, in such cases, it will be necessary to compromise with the ideal and use a non-probability sampling method. This research design decision can be justified on the basis that it is better to have some knowledge, which is restricted because of the type of sample, than to have no knowledge of the topic at all.

However, having said this, it must be stressed that social researchers usually hope that what they find in a sample or group has wider relevance or value (this issue will be taken up in section 8.7 on 'Case Studies'). Therefore, even when non-probability samples are used, they can be selected in such a way that it is possible to make a *judgement* about the extent to which they represent some population or group. What the researcher hopes to achieve is a sound basis for making such a judgement.

Decisions about whether or not to use a probability sample, and how it should be done, are not confined to quantitative studies; they are also necessary in studies that intend to gather qualitative data.⁷ Because qualitative methods are resource-intensive, smaller samples are usually used. Here the compromise is between having data that can be applied to large populations, and having detailed in-depth data on, perhaps, an unrepresentative sample or just a single case. However, arguments can be made that using qualitative methods is not really a compromise, and that these methods can produce a richer understanding of social life than is possible with less in-depth and wider span quantitative methods. The relative merits of qualitative and quantitative methods will be discussed in chapter 9. In the meantime, we need to confine our attention to the role of sampling in both types of data gathering and analysis.

There is no necessary connection between the type of research method used (quantitative or qualitative) and the type of sample that is appropriate. Research using both types of methods can use populations, although they are likely to be much smaller with qualitative methods. For example, it would be possible to study neighbouring behaviour through participant observation in a city street (or block) in a middle-class suburb. Defining the target population as residents of a particular city block may restrict the generalizability of the results, but the richness of the data may allow generalizations, based on a judgement about how typical the chosen block is for that or other cities, or whether other suburbs in other cities are similar in important respects. If the research also included a variety of sites in the same city (e.g. working-class, upper-class, ethnically homogeneous, etc.), then generalizability may be enhanced. Here the sampling issue becomes one of which city, and which streets/blocks, to select. The method of

selection may be *judgemental* rather than being based on probability, although probability sampling should not be overlooked.

Accidental or convenience sampling

The idea of randomness in sample selection should not be confused with selection by accident. The method of accidental or convenience sampling is the most unsatisfactory form of non-probability sampling as it is likely to produce unrepresentative samples. The use of such methods may be an indication of laziness or naivety on the part of the researcher, or may be used in 'quick and dirty' commercial research. A typical convenience sample is obtained when an interviewer stands on a street and selects people for interview *accidentally* as they pass. Such respondents are representative of no particular population, not even of people who passed that spot during a particular period of a particular day. The views obtained in such a study do not even represent the mythical 'person in the street'. Doing accidental interviews at a selection of spots in a city is no help; the population can neither be defined adequately nor its members be given an equal or known chance of selection.

A similar kind of accidental sample is obtained when readers of a newspaper or magazine are asked to complete a short questionnaire that is then cut out and mailed in, or completed on a website. While the population may be defined as the readers of that particular issue of the newspaper or magazine, the self-selection process means that the respondents are representative of nothing. A similar process can occur when a researcher advertises in the print media for people with particular characteristics to volunteer to participate in a study; there is no way of knowing how representative they might be of that population. In some circumstances a researcher may have to use such a sampling method as a last resort, but the results from such a study would need to be heavily qualified.

Quota sampling

This method of sampling is certainly a big improvement on accidental sampling and is commonly practised when it is impossible, difficult or costly to identify members of a population. It has the advantage that it can produce a sample with a similar distribution of characteristics to those that are considered to be important in the population that it is supposed to represent. A set of selection criteria is identified because of their relevance to the research topic, although the establishment of these criteria may not be a simple matter. For example, in a study of undergraduate university students, three selection criteria might be used: gender (male and female), year at university (e.g. first, second, third, and fourth or more); and type of degree being undertaken (e.g. arts, science, engineering, medicine, law, economics or education). These three criteria could produce as many as fifty-six selection categories.⁸ The researcher would then have to decide how many respondents are to be selected in each category. There are two main possibilities: equal numbers in each category or numbers proportional to their incidence in the population (if this is known). The first option can ensure sufficient numbers for analysis, while the second has some

of the advantages of proportional stratified random sampling. However, as the selection into each category will usually be accidental, in neither case can the population characteristics be estimated statistically. In an interview survey, for example, interviewers are only required to fill up the categories with the quota of respondents who meet the selection criteria. The advantages of quota sampling are that it is economical, easy to administer and quick to do in the field. It can also produce adequate results, although the degree to which it does this may not be known. It may even be better than using a probability sample with a poor response rate.

Judgemental or purposive sampling

One use of this method is to deal with situations where it is impossible or very costly to identify a particular population; that is, where there is no available list of the population elements. For example, a study of intravenous drug users could find respondents by a number of means: by contacting users in the field; through police and prisons; or through public and private agencies such as drug rehabilitation centres. Depending on the particular research questions being investigated, it would be possible to contact a significant number of drug users from a variety of contexts, and to include at least the most common types of drug users.

A second use of the judgemental sampling method is for selecting some cases of a particular type. For example, a study of organizational behaviour may use a few cases of organizations that have been particularly successful in achieving what is of interest to a researcher. The selection will be a matter of judgement as to which organizations would be most appropriate. A variation on this would be to select cases that contrast in some way; for example, successful and unsuccessful organizations. Another use would be to select a variety of types of cases for in-depth investigation. For example, a study of 'problem' families could seek the assistance of experts in a social welfare agency to provide a list that includes a variety of such families or, perhaps, families that differ on a set of criteria. The expert, with directions from the researcher, will make a judgement about the appropriateness of families for the study. These judgements may be informed by theoretical considerations.

Snowball sampling

This non-probability method is also known as network, chain referral or reputational sampling. The analogy is of a snowball growing in size as it rolls in the snow. This method has two related uses. In a difficult to identify population, such as intravenous drug users, it may be possible to contact one or two users who can then be asked for contact details of other users, and so on. Another example would be in a study of people who regard themselves as social equals; the respondents' definitions of social equality can be used to build up a sample. Snowballing can also be used to locate natural social networks, such as friendship networks. Once contact is made with one member of the network, that person can be asked to identify other members and their relationships.

Theoretical sampling

Finally, a common sampling method used in some qualitative research is theoretical sampling. When a researcher collects, codes and analyses data in a continuous process, as in grounded theory, decisions about sample size are made progressively. The initial case or cases will be selected according to the theoretical purposes that they serve, and further cases will be added in order to facilitate the development of the emerging theory. As theory development relies on comparison, cases will be added to facilitate this. An important concept in this process is 'theoretical saturation'. Cases are added until no further insights are obtained; until the researcher considers that nothing new is being discovered. Another grounded theory concept related to sampling is 'slices of data', defined as 'different kinds of data [that] give the analyst different views or vantage points from which to understand a category and to develop its properties' (Glaser and Strauss 1967: 65). A variety of slices is desirable to stimulate theory development. Just which slices are selected, and how many, is a matter of judgement. An important point about this method of sampling is that any notion of representativeness is irrelevant.

Accuracy, Precision and Bias

These three important sampling concepts need to be discussed briefly.⁹ They are concerned with the ability of a particular sample, and a particular method of sampling, to be able to estimate a population parameter from a sampling statistic. A *population parameter* is the actual value of a particular characteristic, such as the percentage of females, or the mean age. A *sample statistic* is the value of such characteristics obtained from the sample. The aim, of course, is to draw a sample in which the value of the characteristic is the same as that in the population.

The concept of accuracy refers to the degree to which a particular sample is able to estimate a population parameter. A sample value is inaccurate to the extent that it deviates from the population value. This is referred to as *sampling error*. While it is usually not possible to establish the level of accuracy of an estimate, it is possible to calculate from any one sample value the likely distribution of all possible sample values. This possible distribution indicates the fluctuations in sample values that result from random selection. In other words, the probable accuracy or *precision* can be calculated and used as a basis for estimating the population parameter.

Sampling bias refers to the systematic errors of a particular sampling method. These errors affect the capacity of the method to estimate population parameters. Here we are dealing with not just one sample estimate, but with all possible samples that the method can produce. If it were possible to obtain the value of a sample statistic for all possible samples from a population, and the mean of them was to be compared with the population parameter, the degree of bias of the sampling method could be established. Some of the methods we have discussed are less biased than others; the compromises made against simple random sampling are usually responsible.

There are, therefore, two important considerations in choosing a sampling method. The first is the likely *bias* of the method itself, and the second is the possible *accuracy* of its individual estimates of population parameters. The researcher can deal with the former by selecting a method that minimizes bias and with the latter by making sure that the sample size is appropriate. In general, the larger the sample, the narrower is the distribution of possible sample values, and hence the more precise the estimates. Sample size will be discussed shortly.

Response Rate

Ultimately, the usefulness of any sampling design will be determined by the extent to which all sampled units are included in the study. A poor response rate can destroy all the careful work that has gone into devising an appropriate sampling design. Many years ago a statistician colleague argued that there is no point in trying to estimate population parameters from sample statistics if the response rate is below about 85 per cent. Lower response rates, he suggested, make a nonsense of the application of probability theory because of the lack of precision that can be introduced. Given that this response rate is very rarely achieved, it would appear from research reports that researchers are either unaware of this problem, or just choose to ignore it.¹⁰ Tests of significance are ritualistically applied to sample data that are based on poor response rates, often well below 50 per cent. *To use probability sampling is one thing; to achieve a high response rate is another. Both are needed to estimate population parameters from samples.* It is therefore essential that every effort be made to achieve as high a response rate as possible, whether the study is based on a sample or a population, but particularly with probability samples.

As human populations have become saturated with social surveys and opinion polling, low response rates have become increasingly common. What can a practising social researcher do? The problem with low response rates is the risk of unrepresentativeness. If data are available on population distributions on critical variables (e.g. from a census), sample distributions can be compared with them. If the distributions are similar, then tests of significance can be used with some confidence. When a very poor response rate is anticipated, it may be better to use a carefully designed non-probability quota sample. This will at least ensure that sample distributions will be similar to those in the population, even if representativeness cannot be guaranteed.

Sampling and Tests of Significance

Tests of significance are designed to apply probability theory to sample data in order to draw conclusions as to whether characteristics, differences or relationships found in a sample can be expected to have occurred in the population, other than by chance. Without going into the intricacies of probability theory, what 'chance' means here is the likelihood that one of the possible very 'deviant' samples may have been drawn from the population, thus producing very inaccurate estimates of

population parameters. A test of significance, with a predetermined confidence or probability level (such as 0.05), will estimate the chance that the sample statistic is 'deviant'; that is, the statistic lies outside tolerable limits (the confidence interval). To reiterate, these statistical tests are only relevant when probability sampling has been used, and then only when there is a very good response rate.

In any attempt to generalize from a sample to a population, it is necessary to decide on what is technically called a *level of confidence: the degree to which we want to be sure that the population parameter has been accurately estimated from the sampling statistic*. All such estimates of population parameters have to be made within a *range of values around the sample value, known as the confidence limits*. Just how big this range or interval is depends on the *level of confidence* set. If you want to have a 95 per cent chance of correctly estimating the population parameter, the range will be smaller than if you want to have a 99 per cent chance. For example, if in a sample of 1,105 registered voters, 33 per cent said they would vote for a particular candidate at the next election, we can estimate the percentage in the population to be between 30.2 and 35.8 (a range of 5.6 per cent) at the 95 per cent level of confidence, and between 29.4 and 36.6 (a range of 7.2 per cent) at the 99 per cent level.¹¹ Therefore, *setting the level of confidence high (approaching 100 per cent) will reduce the chance of being wrong but, at the same time, will reduce the accuracy of the estimate* as the confidence limits will have to be wider. The reverse is also true. If narrower confidence limits are desired, the level of confidence will have to be lowered. For example, if you only want to be 80 per cent sure of correctly estimating a population parameter, you can achieve this with very narrow confidence limits; that is, very accurately. Hence, at an 80 per cent level of confidence, the confidence intervals would be between 31.2 per cent and 34.8 per cent (a range of 3.6 per cent). However, this more accurate estimate has limited value, as we cannot be very confident about it. Hence, *there is a need to strike a balance between the risk of making a wrong estimate and the accuracy of the estimate*. (See Blaikie 2003: 171–7.)

Unfortunately, there is no other way of generalizing from a probability sample to a population than to set a level of confidence and estimate the corresponding confidence limits. The commonly used levels of confidence are 95 per cent (0.05 level) or 99 per cent (0.01 level), but these are conventions that are usually used without giving consideration to the consequences for the particular study. It is worth noting again that these *problems of estimation are eliminated if a population is studied*. As no estimates are required no levels of confidence need to be set; the data obtained *are* the population parameters.

The only other way to avoid having to use tests of significance to estimate population parameters from sample statistics, with all its assumptions and risks, is to 'average out' results from many samples from the same population. The possible errors produced by a few deviant samples (and the probability is that there will be only a few if enough samples are drawn) become insignificant. Clearly, this solution is not feasible so we have to content ourselves with tolerating some risk of being wrong in our estimate of a population parameter.

There appears to be a great deal of misunderstanding about the use of tests of significance with populations and samples. It is not uncommon for social researchers to call whatever units they are studying a sample, even when the units

constitute a population. This can lead to the use of certain statistical tests (e.g. the chi-square test for nominal data and the *t* test for interval or ratio data) with population data (parameters) when they are only necessary if sample data (statistics) are being analysed. The fact that these tests are called 'statistical' tests is the clue that they should be applied only to sample 'statistics'.

Tests of significance are applied inappropriately to data from populations or non-probability samples when they are frequently misinterpreted as indicating whether there is any difference or relationship worth considering in the data. Any differences or relationships found in a population are what the data tell you; applying a test of significance is meaningless. The researcher has to decide, on the basis of appropriate measures of difference or association, if the difference or relationship in a population is worthy of consideration. *It is the size of the difference or strength of association between variables, in both samples and populations, not the level of significance,¹² that is relevant to this decision*, and then it is a matter of judgement about whether the relationship is important. If a non-probability sample is used, it is not possible to estimate population characteristics. Hence, the use of tests of significance is inappropriate.

It should now be evident that a critical research design decision is whether to use a population or a sample. This decision will be influenced by the need to strike a compromise between what would be ideal in order to answer the research questions, and what is possible in terms of available resources and other practical considerations, such as accessibility of population elements. The decision will then have a big bearing on the kinds of analysis that will be necessary.

Sample Size

This brings us to the research design question that is asked by students more than any other: 'How big should my sample be?' Of course, the question might mean, 'How big should my population be?' or 'How many people should I have in my research?' There are no easy answers to such questions as many factors have to be considered. In research that uses quantitative data, the answer will vary depending on whether probability or non-probability samples are being used. Some techniques are available to calculate optimum sample size under certain conditions. However, with qualitative data, this is much more difficult. Some attempts have been made to calculate sample size but with limited success. See, for example, Guest et al. (2006), Francis et al. (2010) and the debate in the *International Journal of Social Research Methodology* (Fugard and Potts 2015; Emmel 2015; Hammersley 2015b; Byrne 2015, Blaikie Forthcoming).

Probability sample size

There are four important factors to be considered in deciding the size of probability samples:

- the degree of accuracy that is required, or, to put this differently, the consequences of being wrong in estimating population parameters;