7

Sampling procedures

This chapter

- **outlines the nature and types of sampling**
- **clarifies the rationale and principles of sampling**
- **introduces quantitative and qualitative sampling**
- **elaborates on the meaning and computation of the sample size.**

Key headings

Introduction

One of the most significant issues investigators have to consider when designing a project is the type and number of the people who will be included in the study. In this context, and as noted earlier, researchers will have to take into consideration a number of very important questions, such as:

- Will the whole population or a sample be studied?
- If sampling is preferred, which sampling procedure is most suitable?
- How large should the sample be?
- Is a sampling frame required?
- If so, is one available?

- How representative should the sample be?
- How will possible problems, errors and distortions be prevented?
- What kind of administrative arrangements are required for the selection of the sample?
- Are the required time, funds and staffing available, and if so, how can they be rationally employed?
- How will non-response be dealt with in the study?
- Are there any issues of ethics and objectivity to be considered at this stage, and how will such requirements be met?

The answers to these questions are many and the options diverse. One option is complete coverage of the population (saturation survey), whereby all units of the target population will be studied. In this case the target population is also the survey population. Another option, and the most common, is sampling, whereby the target population is investigated.

Box 7.1

Some basic concepts of sampling

Target population: The population for which information is required.
Survey population: The part of target population that is studied.
Sample: The part of target population that is to be studied.
Sampling: The process employed to extract samples for study.
(Sampling) Units: The persons, groups, systems etc. chosen to be studied.
Saturation survey: A survey that includes all units of the target population.

1 Reasons for sampling

It was mentioned above that sampling enables the researcher to study a relatively small part of the target population, and yet obtain data that are representative of the whole population. Although this is one good reason for researchers to employ this procedure, there are many more that can be put forward in favour of sampling. Those listed in Box 7.2 are the most common (Becker, 1989; Benini, 2000):

Despite these advantages, sampling is argued to be associated with a number of problems, which deserve to be mentioned. The two most obvious problems are, first, that sampling requires more intense and complex administration, planning and programming than saturation surveys; and second, that sampling implies a reduction in the size of the target population and, hence, fewer potential respondents; this raises questions regarding representativeness and generalization of the findings that cannot be ignored.

The response to these logical concerns about sampling is simple and convincing. Sampling may require administration and programming but overall both are less demanding when dealing with a sample (e.g. 5,000 people) than with a whole population (e.g. 5 million). Finally, samples are chosen in such a way that the demand for representativeness and generalization is not compromised. In this sense, studying fewer people can be a strength and not a weakness.

Box 7.2

Why use samples?

- *Necessity*. In many cases a complete coverage of the population is not possible.
- *Effectiveness*. Complete coverage may not offer substantial advantage over a sample survey. On the contrary, it is argued that sampling provides a better option since it addresses the survey population in a short period of time and produces comparable and equally valid results.
- *Economy of time*. Studies based on samples take less time and produce quick answers.
- *Economy of labour*. Sampling is less demanding in terms of labour requirements, since it covers only a small portion of the target population.
- *Overall economy*. It is also thought to be more economical, since it involves fewer people and requires less printed material, fewer general costs (travelling, accommodation etc.) and of course fewer experts.
- *More detailed information*. Samples are thought to offer more detailed information and a high degree of accuracy because they deal with relatively small numbers of units.

2 Principles of sampling

Samples are expected to be representative. To achieve representativeness, sampling procedures are expected to follow certain standards and methodological principles. Particularly with regard to quantitative research, researchers and writers alike consider those listed in Box 7.3 to be very important (see Pfeifer, 2000; Selltiz et al., 1976).

Additional or alternative standards and principles will be introduced when considering specific types of sampling, especially when dealing with sampling in qualitative research. Researchers who work outside quantitative parameters have raised questions about sampling, in particular about objectivity and the rigid formulation of its principles, but this does not detract from the validity of these elements.

Box 7.3

Principles of sampling

- Sample units must be chosen in a systematic and objective manner.
- Sample units must be easily identifiable and clearly defined.
- Sample units must be independent of each other, uniform and of the same size, and should appear only once in the population.
- Sample units are not interchangeable; the same units should be used throughout the study.
- Once selected, units cannot be discarded.
- The selection process should be based on sound criteria and should avoid errors, bias and distortions.
- Researchers should adhere to the principles of research (discussed in Chapter 4).

3 Types of sampling

Sampling procedures vary considerably. Samples may be constructed through self-selection (respondents decide to take part in a study, for example, in response to media calls for volunteers) or, as is most common, through the researcher. There are also sampling procedures based on probability standards (random or probability samples), and on non-probability standards (non-probability samples). In the following sections we shall discuss probability and non-probability sampling procedures. The main criteria for these samples are listed in Box 7.4.

3.1 Probability (random) sampling

Probability sampling is the procedure in which the choice of respondents is guided by the probability principle, according to which every unit of the target population has an equal, calculable and non-zero probability of being included in the sample. There are several forms of probability sampling, but simple and systematic random sampling are the most common. These two types of probability sampling and the methods they employ will be considered below.

3.1.1 Simple random sampling

Simple random sampling is marked by the fact that the sampling units have an equal chance of being selected and are independent from each other. Their chance of being selected does not depend on the selection of other units. The three most common methods of simple random sampling are: the lottery method, the random numbers method and the computer method.

Box 7.4 **Criteria of probability and non-probability samples**

Probability sampling:	Non-probability sampling:
Employs probability theory	Does not employ probability theory
Is relatively large	Is small; often including a few typical cases
Its size is statistically determined	Its size is not determined statistically
Its size is fixed	Its size is flexible, but can also be fixed
Is chosen before the research	Is chosen before and during the research
Controls researcher bias	Does not control researcher bias
Involves complex procedures	Involves simple procedures
Its parameters are fixed	Its parameters are flexible
Involves high costs	Involves relatively low costs
Planning is time consuming	Planning is not time consuming
Is aimed to be representative	Representativeness is limited
Planning is laborious	Planning is relatively easy
Treats respondents as units	Treats respondents as persons
Facilitates inductive generalizations	Facilitates analytical generalizations
Is employed in quantitative research	Is mostly for qualitative researchers

The lottery method

The lottery method procedure is conducted following the four steps shown below.

Step 1: Identify or construct a sampling frame, that is, a list of the units of the target population. Such frames may for instance be the electoral role, student records, rating records or similar lists. Include the names and addresses of all sample units in alphabetical order, numbered accordingly.

Step 2: Allocate the sample size, that is, the number of units required for the study.

Step 3: Place a number of small discs or balls in a container, numbered to correspond to the names contained in the sampling frame. If 500 names are listed in the frame, there should be 500 balls or discs in the urn, numbered from 1 to 500.

Step 4: Mix well and remove one ball from the urn. The number of this ball is registered and the corresponding name in the sampling frame is noted. This is the name of the first respondent. The ball is either returned to the urn or left out; either method may be employed. Continue this process until sufficient names have been selected. (If an already drawn number is selected it is ignored). The selected respondents constitute the sample.

The random numbers method

This method is similar to the lottery method, except that the urn and balls are replaced by random number tables, which are available in separate publications or in the appendices of statistics texts. Choosing the sample by using the random numbers method involves the following steps:

Step 1 Identifying or constructing a sampling frame.

Step 2 Selecting appropriate tables of random numbers.

Step 3 Picking numbers from the tables randomly and registering them; the names in the sample frame that correspond to these numbers constitute the sample.

The computer method

In this method, we enter the list of (say 6,000 numbered) sample units in the computer and then instruct the computer to select a set of numbers equal to the number of sample units, for example, 500 numbers ranging between 1 and 6,000. Following this procedure 500 numbers will be produced, or even the numbered names of the selected units will be produced, depending of course on the computer software. This list is the sample.

> *Example C*. The sample for the 'students and feminism' study is to be selected by the computer method. We proceed as follows:
> 1. The appropriate sampling frame is obtained.
> 2. The sample size is determined; this is 500.
> 3. The computer is instructed to randomly select 500 numbers from 1 to 6,000.
> 4. The numbers are identified as in the previous examples.
> 5. These numbered names constitute the sample for the study.

Although all three sampling methods are efficient and commonly used, the electronic construction of samples has become more popular than in the past due to increased computerization and accessibility of public records and electronic telephone books; more-over, computers have become an integral part of the research context. Where appropriate records are available and suitable, the computer can choose not only numbers but also names, addresses, telephone numbers and other contact details, and can even contact the respondents on request. Drawing numbers from the urn and similar techniques have become methods of the distant past.

3.1.2 Systematic random sampling

Systematic random sampling is a procedure in which the sampling units are not only chosen randomly – as in simple random sampling – but in which this random choice is also integrated with the choice of another sampling unit. This is what distinguishes this type of sampling from simple random sampling, and what qualifies it to be termed systematic. The actual choice of units is orchestrated through a system of computation that aims, first, to maintain randomness in selection, and second, to spread the sampling units evenly throughout the list of respondents (the sampling frame). The system is based on the *sampling fraction method*.

In this method, units are drawn from a sampling frame by means of the sampling fraction (symbolized by k) that is equal to N/n, where N is the number of units in the target population and n the number of units of the sample. For instance, if the target population is 4,800 and the intended sample size 600, the sampling fraction is 8 (k = 4,800/600 = 8). To select a sample by using the sampling fraction method, we proceed as follows:

Step 1: Identify or construct a sampling frame.
Step 2: Determine the sample size.
Step 3: Compute the sampling fraction k (as above, $k = N/n$).
Step 4: Randomly select a number between 1 and k. In the above example, since $k = 8$, the random number would be between 1 and 8; let us say 6.
Step 5: Record the random number (6) and every eighth number after 6, until 4,800 is reached, e.g. 6, 14 (6 + 8 = 14), 22 (14 + 8 = 22), 30 (22 + 8 = 30) etc.
Step 6: Locate the names in the sampling frame that correspond to the selected numbers.

The respondents thus identified constitute the sample.

> *Example D*. The local Telecom office wants to survey the 11,000 customers who joined a half-price programme six months ago and assess their degree of satisfaction with this programme. We are to conduct this survey, and in the first instance we need to draw up a sample using the systematic random sampling procedure. The intended sample size is 500. A list of the customers in question is available. How do we choose the 500 respondents?
>
> 1. Obtain the sampling frame from the Telecom office.
> 2. Compute the k fraction; this is 11,000/500 = 22.
> 3. Randomly select a number between 1 and 22; in this case, 18.

4. Compute the numbers for each of the 500 respondents by progressively adding 22 to 18 and to the resulting sums. The numbers will be 18, 40 (18+22 = 40), 62 (40+22 = 62), 84 (62+22 = 84) and so on.
5. Locate the names in the sampling frame that correspond to these numbers. These are the customers who will be included in the sample.

It goes without saying that more recent computer software can assist the researcher to conduct this procedure without any manual intervention on the part of the researcher. The computer constructs and presents the sample list in no time at all.

3.1.3 Stratified random sampling

Stratified random sampling is a *probability sampling procedure* in which the target population is divided into a number of strata, and a sample is drawn from each stratum. The resulting sub-samples make up the final sample of the study. The strength of this procedure is in that it allows all population groups to be represented in the final sample. The division of the population into strata is based on one or more significant criteria, such as sex, age, ethnic background, race or economic status, but more so on criteria that are thought to be associated with the research topic.

The sample size can be proportionate or disproportionate to the units of the target population. This means that the samples taken from each stratum can be either proportional or disproportional to the size of the samples. As indicated above, a stratified sample is employed when there is a need for representing all groups of the target population in the sample, and when the researcher has a special interest in certain strata. In this sense, the method is very economical, and offers a high degree of representativeness. A stratified sample is drawn as follows:

Step 1: The target population is divided into a number of strata, according to the number of the significant groups in the population.
Step 2: The sampling frames for each of these groups are identified; if these are not available, relevant frames must be developed.
Step 3: Employing one of the methods discussed above, a sample is drawn from each group. This can be proportionate or disproportionate to the number of units in the population.
Step 4: The individual samples are merged into one; this constitutes the sample for the study.

Example E. We are asked to study the attitudes of our community to the government's foreign aid policy. Stratified sampling has been chosen in order to include in the study all ethnic groups. To obtain the names of the respondents, we proceed as follows:

1. Sampling frames are identified or prepared for each ethnic group in the community.
2. A decision on whether a proportionate or non-proportionate stratified sampling should be employed is made.

3. The number of subjects to be chosen from each ethnic group is determined. This is 60 per cent for Asians, 20 per cent for Africans, 15 per cent for Greeks and 5 per cent for Germans.
4. One of the methods discussed above (e.g. the lottery method, random numbers, sampling fraction) is used to choose the sub-samples (i.e. the separate samples from each ethnic group).
5. The sub-samples are merged into one sample.

The resulting respondents constitute the final sample. In this way, the researcher can expect a relative representativeness of the major ethnic groups in the study, which is, as would be expected, larger than if the target population was not stratified. As noted above, computers can assist by processing and producing the required lists without manual intervention and in no time.

3.2 Multi-stage sampling

In multi-stage sampling, the selection of sample units begins with the choice of a large sample, and *proceeds* with new samples taken in succession from those previously selected, thus facilitating the construction of a more suitable and more effective choice. More specifically, a large sample is chosen, using a random sampling procedure, and then another sample is taken from within this sample, excluding excess and unrelated units. For instance, if the study is to focus on professional women, all men and non-professional women contained in the first sample will be discarded. If required, another sample is chosen from the second sample for similar reasons; for example to exclude professional but retired women from the sample. This process is continued for as long as required, with each additional drawing making the sample more specific, more focused, more relevant to the research object and more representative. The characteristic of this type of sampling is that data collection is conducted only from the final sample.

The process of choosing a sample through the multi-stage sampling method proceeds as follows (see Pfeifer, 2000):

Step 1: A sampling frame for the target population is identified.
Step 2: A large probability sample is chosen; the units of this sample are usually referred to as primary selection units. A sample from the primary selection units is then chosen.
Step 3: After the criteria of the respondents have been identified (in terms of gender, ethnicity, marital status etc.), another sample is drawn from within this sample. In most cases, a second sample is sufficient to meet the requirements of the study. Otherwise, the procedure is repeated until the targeted sample size is reached.
Step 4: The final group constitutes the sample of the study.

Example F. We are to study the attitudes of nurses to the proposal that compulsory reporting of suspected domestic violence injuries should be abolished. The sample is to be determined by means of multi-stage sampling procedures. To obtain the sample we proceed as follows:

1. A list of nurses from all hospitals in the region is obtained.
2. From this list, 2,000 nurses are chosen.
3. Using hospital unit as criterion, 600 nurses are chosen, representing all major hospitals of the country.
4. These 600 nurses are finally screened, first using gender as a criterion, and then specialisation, age, and length of employment, resulting in a balanced sample of 150 respondents. A non-probability method is used.
5. These respondents constitute the final research sample.

The use of several screenings and drawings is not only time-consuming but also expensive. Hence, multi-stage sampling procedures are employed only when absolutely necessary. An obvious case for such a sampling procedure is a heterogeneous population where there is not enough information to permit the construction of a representative sample without screening. Again, computers can provide the same results in less time, if the appropriate software are available.

3.3 Cluster sampling

Cluster sampling is a procedure in which the researcher *chooses the study units progressively*, beginning with clusters and moving to smaller groups within them, before the final sampling units are considered. This sampling method is employed primarily when a sampling frame is either unsuitable or not available. Cluster sampling is also employed when alternative methods are too expensive, and particularly when clusters are an important research factor. Such clusters include schools, classes, soccer teams, hospitals, small business and other well-integrated groups with a common identity. The following is an example.

Example G. We are interested in the attitude of technical-school teachers to a recent government report on subsidies to these educational institutions. The study is to be conducted by means of a survey. Cluster sampling will be used in order to allow schools affected by the new proposed measures to respond. To compile the sample we proceed as follows:

1. A list of all technical schools is constructed (sampling frame), or obtained from the relevant authorities if available (and if accessible).
2. The number of schools required for the study is determined by employing appropriate standards.
3. Schools are chosen from the sampling frame by means of one of the sampling methods introduced above (the lottery method, sampling fraction etc.).
4. The number of teachers required for the study is determined and then divided by the number of schools to determine the number of teachers to be chosen from each school (proportionate or disproportionate numbers to be considered).

5. The required number of respondents from each school are chosen, using one of the methods introduced above.
6. The sum of all teachers chosen this way constitutes the sample.

It should be noted that, although it is clearly a distinct technique, cluster sampling operates in a manner similar to that of stratified sampling and, even more so, of multi-stage sampling. Some writers employ this method to choose subjects spread in large areas, as with area sampling. In either case, this sampling procedure has its own identity in dealing with clusters, and is systematically employed when conditions require it. Again, computers can provide the same results in less time, if the appropriate software are available.

3.4 Multi-phase sampling

The sample selection within this procedure is identical to multi-stage sampling, with the difference that in this sampling procedure, *each sample is adequately studied before the next sample is drawn*. This offers an advantage over other methods, because the information gathered at each phase helps the researcher to focus the selection more effectively and more constructively in later phases.

> *Example H*. A social worker is interested in the division of labour among homosexual couples. The sampling procedure is to be multi-phase sampling. To compile the sample we proceed as follows:
> 1. A sampling frame of homosexual people is constructed, and 500 are randomly chosen.
> 2. These respondents are interviewed about the nature of their relationship, their age, and other demographic criteria.
> 3. Of these respondents, 300 (150 gay and 150 lesbian couples) are chosen randomly who have a permanent partner and at least one child living in the same household.
> 4. This stratified sample will be the final sample to study the division of labour among homosexual couples.

You will have noted that the use of multi-phase samples is employed very constructively and effectively by researchers to facilitate research where it is not otherwise possible, and to achieve what single-sample studies cannot.

3.4.1 Area sampling

Area sampling is a procedure in which *multi-stage sampling is applied in geographical areas*. More specifically, the samples are chosen as follows:

Step 1: Framing the area.
Step 2: Dividing it into large parts.
Step 3: Randomly selecting a number of representative parts.
Step 4: Dividing the selected parts into units that are small enough to be studied directly.

Step 5: Drawing a representative sample of units from each part.

This will be the sample for the study. A typical example is given below.

> *Example I*. A journalist is interested in the views of people, living in country towns, on a proposed Euthanasia Bill. The proposed sampling procedure is to be area sampling. In this typical case, sampling proceeds as follows:
> 1. A sampling frame of country towns is established, and a sample of five towns is drawn.
> 2. The chosen towns are divided into suburbs, and four suburbs from each of the five towns are chosen randomly, resulting in 20 suburbs to be studied.
> 3. The streets in each chosen suburb are listed and a sample of five streets is chosen from each. This gives 100 streets registered for consideration.
> 4. The households in each of the chosen streets are now listed, and ten households are chosen from each street for study. This gives 1,000 households.
> 5. The sample will include the heads of all households chosen in the final draw.

This sampling procedure is employed when the researcher has reasons to believe that other sampling procedures may not allow all geographical areas to be represented in the study.

3.5 Spatial sampling

This procedure is employed when the study *addresses people temporarily congregated in a space*, and the data have to be collected before the crowd is dispersed. An example of such cases is the study of the views of people demonstrating in a city square about tax policies. Due to the nature of the population, there are neither sampling frames nor sufficient time available to permit the use of other methods. Apart from this, data collection has to be conducted so that a relatively representative coverage is achieved, randomly and in a systematic way before the crowd disperses. The way this is usually done is shown in the following example.

> *Example J*. In a sit-in at the local university, students have occupied the main administration building and refuse to leave. The Department of Sociology quickly decides to investigate certain aspects of this demonstration, including the reasons for participating in it, and the type of students joining such a demonstration. The choice of the respondents proceeds as follows:
> 1. Ten interviewers line up at the front of the room where students are gathered.
> 2. As instructed, the interviewers address the person who happens to be in front of them, and ask the study questions.
> 3. Then, they all move five steps forward and approach the person who is now in front of them.
> 4. They proceed further, in the same way, until they reach the back of the room.

5. The students interviewed constitute the sample for the study.

Obviously, the details of the procedure can be changed to meet the actual circumstances of the situation. For instance, if a crowd is large, the interviewer might have to walk a longer distance, say ten steps. Also, it may be necessary for the interviewer to seek specific people; for example, the first will be male, the second female, and so on, or to include a variety of respondents (young, old, ethnic, non-ethnics and so on).

4 Non-probability sampling

As the name indicates, non-probability sampling procedures do not employ the rules of probability, do not ensure representativeness, and are mostly used in exploratory research and qualitative analysis. Some of these techniques can, with some adjustment, be converted into probability methods. Accidental sampling, purposive sampling, quota sampling and snowball sampling are examples of non-probability sampling techniques; they are presented below.

4.1 Accidental sampling

This procedure employs no systematic techniques to choose the respondents. Instead the sample units are *those people who 'accidentally' come into contact with the researcher*. For instance, the researcher may stand at a street-corner, in front of a school or church, or at the main exit of a shopping centre, and ask a number of people passing by to take part in the study. They are chosen 'by accident' – they just happen to be there at that time – hence the name of the sampling procedure (there are several other names for it, including 'convenience sampling', 'chunk sampling', 'grab sampling' and 'haphazard sampling'). An example is given below.

> *Example K*. The local chamber of commerce in a small country town want to study people's reasons for shopping in the four large supermarkets. The sampling procedure is accidental sampling. The study proceeds as follows:
> 1. Two interviewers stand at the door of each supermarket with instructions to address 100 shoppers passing by with relatively full supermarket trolleys, asking them the relevant questions.
> 2. The completed forms are returned to the researcher on the same day for evaluation.

The researcher here is not interested in representativeness, objectivity, validity or similar considerations, but in getting information that would reveal certain aspects of the lifestyle in question and in certain cases give information about typical cases.

4.2 Purposive sampling

In this technique the researchers purposely *choose subjects who, in their opinion, are relevant to the project*. The choice of respondents is guided by the judgement of the investigator. For this reason it is also known as judgemental sampling. There are no particular procedures involved in the actual choice of subjects. Here is an example:

> *Example L*. A researcher is interested in the problems of immigrants in a particular community. To explore this topic, it is decided to interview key informants such as the local priests, club secretaries and functionaries of ethnic welfare groups. In the investigator's view, these people can offer adequate and useful information that will give a picture of the problems facing immigrants.

In such cases the important criterion of choice is the knowledge and expertise of the respondents, and hence their suitability for the study.

4.3 Quota sampling

Quota sampling is a procedure in which *the researcher sets a 'quota' of respondents to be chosen from specific population groups*, defining the basis of choice (gender, marital status, ethnicity, education etc.) and determining its size (e.g. 60 parents of toddlers; 35 policewomen; 66 teachers and so on). The choice of the actual respondents is usually left up to the interviewer.

More specifically, the researcher considers all significant dimensions of the population and ensures that each dimension will be represented in the sample. This is usually referred to as dimensional sampling and is particularly useful when the sample is small. In such cases, this procedure guarantees that at least one case from each dimension of the population will be included in the sample.

> *Example M*. The Health Commission is interested in identifying the state of health of workers employed in mining industries around the country. Instead of going through the process of compiling sampling frames in each industry, and then choosing the respondents, the researcher decides to use quota sampling. This proceeds as follows:
>
> 1. Interviewers are sent to each major mining industry.
> 2. The interviewers are told to study, in each unit: 10 workers aged below 20; 10 workers aged 21–30; 10 aged 31–40; 10 aged 41–50; 10 aged 51–60; and 10 aged over 60.
> 3. The interviewers are also told to consider length of service in their selection. Each of the above quotas must contain the same proportions of workers employed in these industries for more and for less than two years.
> 4. The persons who took part in the study constitute the sample.

Quota sampling is quite common in the social sciences because it is less costly than other techniques, does not require sampling frames, is relatively effective, and can be completed in a very short period of time, without jeopardizing the quality of sampling. It is limited, however, especially with respect to representativeness and control of sampling and fieldwork requirements, which in such studies are not relevant. It must be noted, however, that the choice of the respondents can be determined more strictly by employing probability rules, hence requiring sampling frames and specific methods of selection, while retaining the quota factor.

4.4 Snowball sampling

In this approach, the researcher *chooses a few respondents*, using accidental sampling or any other method, and *asks them to recommend other people* who meet the criteria of the research and who might be willing to participate in the project. This process is continued with the new respondents until saturation – that is, until no more substantial information can be acquired through additional respondents – or until no more respondents are available.

> *Example N*. The author (Sarantakos, 2000b) was interested in the lifestyle of same-sex couples and wanted to investigate the establishment of such relationships, as well as their structure, process, stability and quality of life. Given that there were no records on homosexual couples to be used as sampling frames, snowball sampling was used. The sampling design was as follows:
>
> 1. A number of same-sex couples identified through accidental sampling in Australia, New Zealand, Austria and Germany were located and interviewed.
> 2. These respondents were interviewed and then asked to recommend further homosexual couples who could take part in the study.
> 3. After these new couples were interviewed, inquiries for further same-sex couples were made, and additional respondents were secured.
> 4. This process was continued until the sample was saturated.

This method is employed when the lack of sampling frames makes it impossible for the researcher to achieve a probability sample, when the target population is unknown, or when it is difficult to approach the respondents in any other way. In many cases, snowball sampling is the only way of securing a sample for a study.

4.5 Theoretical sampling

In the words of its creators (Glazer and Strauss, 1967: 45), theoretical sampling is the '*process of data collection for generating theory*'. The focus here is on data collection rather than on the choice of respondents. Put simply, in theoretical sampling the sample units are not 'chosen' by the researcher prior to the commencement of the study but during the study, guided by the knowledge that emerges during the study (Burgess, 1984; Strauss, 1991). The researcher chooses the first respondent, collects relevant information and gains knowledge about the research topic, and on the basis of this decides which person is suitable to be interviewed next. The direction of 'theory' that emerges during the research process determines who the next respondent will be, a decision that could have not been possible at the start of the study. Theoretical sampling is interconnected with data gathering and serves to enable comparisons in time and place so as to discover variations in concepts and to integrate categories in terms of their properties and dimensions (Strauss and Corbin, 1998: 201).

Box 7.5

The nature of theoretical sampling

- Entails an ongoing process. The sample here is not chosen and fixed at the outset and before the commencement of research, but chosen in an ongoing process that goes right through the whole study.
- Involves 'places, people and events'. Theoretical sampling takes into account the fact that people think and act differently, depending on many factors. Taking these factors into consideration helps the researcher to test, verify and contrast emerging concepts, categories and theory.
- Is guided by the emerging theory. Theoretical sampling is self-regulated in the sense that it guides data collection and analysis towards developing a theory, which in turn directs the nature and content of sampling.
- Is concerned with developing and validating theory. It is geared towards assuring that the emerging theory is adequately tested, so that it can be granted validity and high quality.
- Ends when theoretical saturation has been reached. It is not the identification of the respondents but the completion of the research that brings sampling to an end.

(Benini, 2000)

In theoretical sampling, the study does not continue until all respondents have been interviewed but rather until the process of study indicates that saturation has been achieved; that is when data collection no longer generates new data, when the categories are 'well developed in their properties and dimensions', and when the relationships among these categories are 'established and validated' (Strauss and Corbin, 1998: 212).

Example O. Assume we are interested in establishing some firm ideas about the coping strategies employed by a group of people who have lost their jobs as a result of a factory closure on the outskirts of the city. We are also interested in how they perceive and experience their new roles and their lives in their family and community. How do we go about choosing the unemployed people? A very elementary description of the relevant parts of the study is given below.

1. Approach one of the people affected by the closure of the factory and address the research topic.
2. Analyse the collected data and, while working on the main issue of the research, look for hints indicating that new aspects of the issue and additional information are available, and which people can provide such information.
3. Approach such a person. Continue the interview with the new person, explore the new facet of the issue and search for further information that would contribute to theory development and indications of where such

information is available (i.e. what kind of person could provide such data). Find and contact such a respondent.

4. This process of gaining new information and new informants will continue until saturation is achieved (i.e. until no more new information is found on the research subject).

You may have noticed that this procedure resembles that of snowball sampling, except that here the hints as to new respondents are not given directly by the previous respondent but by the new information and the demands of the developing theory. This process will be understood better when the data analysis of grounded theory research is presented later in this text.

5 Sampling procedures in qualitative research

Qualitative researchers employ sampling procedures that correspond to the philosophy of this type of research, and that are less structured, less quantitative and less strict than the techniques quantitative researchers employ (Coyne, 1997; Higginbottom, 2004). Normally, qualitative studies employ a form of non-probability sampling, such as accidental or purposive sampling (Kuzel, 1992), as well as snowball sampling and theoretical sampling.

Qualitative sampling is guided by the nature of the underlying qualitative framework, which is perceived as an investigative process, not very different from detective work, where 'one makes gradual sense of a social phenomenon, and does it in large part by contrasting, comparing, replicating, cataloguing and classifying the objects of one's study' (Puris, 1995). Sampling here comes after factors and conditions become clear and directive; making decisions about sampling before the study has begun is neither proper nor useful.

Nevertheless, qualitative research has no strict, agreed rules for sampling employed by all researchers. Sampling procedures employed by qualitative researchers include those mentioned above (accidental, purposive, snowball sampling and so on) or a version or combination of quantitative sampling procedures. In all cases, sampling is closely associated with theory. It is therefore either theory-driven 'up-front', where subjects are chosen before data collection guided by theory, or progressively, where they are chosen during data collection. The latter is known as *theoretical sampling*, and is connected with grounded theory (Koerber and McMichael, 2008).

Irrespective of the type of sampling chosen, several sampling parameters must be considered before a qualitative study can begin. Although qualitative sampling is a function of the research process itself that is decided on while the research is in progress and depends on the outcome of the study, researchers do have to decide at the outset about a number of issues. Such issues are the informants or respondents who will be studied, the setting where research will take place, the events and processes to be considered in the investigation, and the time when research will be conducted. In any case, sampling procedures in qualitative research are inevitably related to a number of issues and choices, a few of which are listed below.

- *Kind of people*: The kind of people (actors) who will be included in their study, i.e. those to begin the study with.
- *Time*: The study may be conducted on working days, on weekends, during the school holidays, in summer or winter, in the afternoons, in the evenings or at any other time.
- *Kind of event*: The kind of event or processes to be studied; whether it will be a routine event, a special event, an unexpected event, or all types of events.
- *Setting*: The context in which the research will be conducted (e.g. the home, the club, the work place, a friend's house).

In summary, the sampling procedures employed by qualitative researchers demonstrate a number of characteristics; those presented in Box 7.6 are considered by a number of writers (see Tuckett, 2004; Berger et al., 1989; Lamnek, 1988) to be most significant.

Box 7.6

Criteria of qualitative sampling

Qualitative sampling is directed:

- not towards large numbers of respondents but rather towards typical cases
- not towards fixed samples but towards ones that are flexible in size, type or subjects
- not towards statistical or random sampling but towards purposive sampling
- not towards 'mechanical' sampling but towards theoretical sampling
- towards fewer global settings than quantitative sampling
- not towards choosing a sample before the study has started, but (often) while the study is in progress
- not towards a strictly defined size but a sample whose number will be adjusted while the study is in operation
- not towards representativeness but rather towards suitability.

As stated earlier, the range of qualitative research has significantly widened in scope and purpose during the past 20 years, offering researchers more choice than before. It is interesting, then, to note that a trend is gradually becoming more evident within certain areas of qualitative research, whereby the sampling repertoire includes a wide range of methods and procedures, including tools such as *a priori* determination, complete collection, extreme case sampling, typical case sampling, maximal variation sampling, intensity sampling, critical case sampling, sensitive case sampling, convenience sampling, primary selection and secondary selection (Flick, 1998). Obviously, sampling is equally valuable in both research domains; only their structure and purpose are different.

6 Sample size

The question about appropriate sample size in social research has attracted due attention from researchers of all schools of thought (Krämer and Thieman, 1987). However, the focus of relevant estimations varies significantly, with some showing an interest in pure quantity, others in quality and others again in both. A wise rule in this case is: the sample must be 'as large as necessary, and as small as possible'. This critical figure is reached in some cases through logical estimates, and in others through statistical computation, as we shall see next.

6.1 Non-statistical estimations

Sample size is directly associated with two major factors: the paradigm that guides the research, and the nature of the target population. These are the major determinants of the size of the sample, at least in logical terms. In quantitative research, both are seriously considered when the sample size is addressed. In qualitative research, the paradigm guides the process, but the nature of the data obtained will determine the size, and this is unpredictable. The study will stop when saturation is achieved, and this emerges out of the data and not out of logical thinking or other calculations. There are, however, qualitative researchers who follow the quantitative paradigm, and estimate their sample size in advance. Hence, the matter is not that simple!

Box 7.7

Some determinants of sample size:

- *Underlying methodology*: Quantitative research requires larger samples than qualitative.
- *Nature of the study object*: Some research topics require large and others small samples.
- *Available time and resources*: In quantitative research, the more time and resources are available the larger the sample size can be.
- *Homogeneity of the target population*: The more homogeneous the target population the smaller the sample can be, and vice versa.
- *Accuracy*: The higher the degree of accuracy required, the larger the sample.
- *Nature of the data required*: If quantitative data are required, a large sample is needed; if qualitative data are required, a small sample will be sufficient.
- *Purpose of the study*: If the study aims to achieve inductive generalizations, a large sample will be required.
- *Intensity of the study*: The more intense and in-depth the method of data collection, the smaller the sample size.
- *Nature of the study*: Surveys require a large sample; case studies do not.
- *Response rate*: The lower the expected response rate, the larger the sample size.

6.2 Statistical computations

Many quantitative researchers employ statistical methods in order to define the 'right' size of the sample. This is based on the assumption that if certain data are available, the sample size can be statistically computed so that sampling errors can be reduced to a minimum or to an acceptable or expected level. There are several methods employed by statisticians and social researchers, some of which are quite complicated and beyond the limits of this treatise.

In general, the logic of many statistical methods relates sampling error to the standard error (SE): if the standard error is reduced, the sampling error is also reduced. The standard error depends on the size of the sample: with increasing sample size the standard error is decreased. Thus an acceptable standard error can be achieved by changing the sample size. This method manipulates the size of a sample by increasing or reducing it, until it corresponds to a standard error that is considered acceptable. This is then the ideal sample size.

The standard error varies inversely with the square root of the sample. If, for instance, we intend to halve the standard error, we have to quadruple the sample size. Thus, if we wish to determine the sample size that will reduce sampling errors, we start with a sample size taken at random, compute the standard error and increase the sample size until the relevant standard error is at an acceptable level. This works well with small samples, where increases in sample size result in increases in accuracy (i.e. decreases in sampling error), but it does not work equally well with large samples. Above a certain point, the increase in size required to achieve a significant decrease in the error is so large and therefore so costly that it makes such an increase not worth the effort.

In the main, the method employed depends on whether the estimation is directed towards means or proportions. In the former, investigators are interested in ascertaining trends and average scores in the area of study. For example, how large should my sample be to study the average amount of money spent by female students on alcohol per year? In the latter, researchers endeavour to estimate the proportion of people acting in a certain way. For instance, what is the proportion of female students who support affirmative action in this country? There are statistical techniques available to estimate the relevant sample size in both cases.

6.3 Estimating sample size through tables

There is an easier way of estimating the 'right' sample size, without needing to use formulae and computations. This is possible by means of tables. The researcher who wishes to know how large the sample should be needs only to look at the table and, considering the necessary factors, such as p, q, Z and E, computes the figure that corresponds to the required sample size. Simply, p is the population estimate, q is the difference between p and 100, Z is the chosen confidence level, and E the maximum deviation from the true proportions considered by the researcher as acceptable.

Such tables are many and diverse and are constructed as shown above, that is, by means of relevant formulae, this time for every possible combination of p, q, E and z values. The more factors considered and the more detailed the values of these factors, the more detailed and accurate the table and, unfortunately, the more cumbersome the identification of the sample size figure. In tables with many factors and factor values, for instance, one has to decide on the right factors and values, search for the right column and row, and determine the required figure. When there are fewer factors, the tables are easier to use but offer limited information. Such tables may, for instance, offer advice about a certain p/q combination, one confidence level (95 per cent or 99 per cent) and one option for degree of accuracy.

Parten (1950: 314–15), for instance, published in 1950 two tables, one for the 0.05 confidence level and one for the 0.01 confidence level. The tables offer sample size estimates for dichotomous population percentages (p and q) for two levels of confidence and for error limits ranging from 0.25 to 10. They provide useful information and save time and energy when deciding the required sample size.

Krejcie and Morgan (1970) offered an even easier table for estimating sample size. The only information needed to estimate sample size is the size of population. Consequently, their table gives figures for populations ranging from 10 to 1,000,000 people and the corresponding figures for the required sample size. This table computes the sample size by means of another formula, which takes into consideration chi-square for 1 degree of freedom, the population size, the population proportion, which is set at 0.50, and the degree of accuracy, which is set at 0.05, by using a formula developed by the research division of the National Education Association (USA), published in 1960.

Estimating the required sample size by using this table is just a matter of reading numbers! For example, to specify the required sample size you need to know the size of the target population. For a population of 260 people the suggested sample size is 155; for a population of 1,600 the sample size is 310, and for a population of 20,000 the sample size is 377.

A word of caution: the statistical procedures presented above provide a handy tool for estimating the sample size required in each case when a study is to be carried out and when the sample is to be representative. Although these procedures are statistically sound, they relate to estimations based on assumptions and conditions, on which the estimates depend. As shown above, the sample size depends on the values of p, Z and E, chi-square and so on – values that are often difficult to estimate. How can we be sure, for instance, that p is 10, 25 or 48, or that σ is 50, particularly when our knowledge of the population is restricted? Guessing the value of the standard deviation of a population for which we intend to estimate the unknown mean is again a daring estimation. Figures obtained through these procedures must be interpreted within their boundaries.

7 Internet sampling

The popularity of and easy access to the Internet has impacted on the conduct of social research in many ways. This is clearly shown in sampling, which has begun to adjust its techniques to the changing ways of approaching people in the community. As a result of this, Internet sampling has become a part of the research armoury of modern researchers, with increasingly many research bodies using it as their preferred sampling procedure.

In simple terms, Internet sampling is a sampling procedure that is administered, partly or fully, through the Internet. This entails procedures which enable the researcher to bring questionnaires to the attention of prospective researchers, by either directly forwarding them the questionnaire, or informing them of the availability of the survey and asking them to participate. This is facilitated through email or through web pages.

- *Email:* Researchers who gain access to email lists act in one of two ways. They might send a message to the address, asking the email holder to volunteer and take part in the survey. This is usually conducted as spamming, the uninvited email. Otherwise, a researcher can attach the questionnaire to an email sent to all members of an email list, with a note in which they invite them to participate. In a number of cases, participation is associated with a variety of rewards.

- *Web pages (URL):* The same procedure is employed when respondents are sought through a web page, where readers are either asked to complete a questionnaire, or if they agree are directed to the questionnaire. Internet users will come across the researcher's message or the commercial of an advertising agency, or both, and will respond according to their interests.

In a sense, Internet recruiting of respondents is not very different from advertising a study in the media, or accessing people on the phone. Like the latter, Internet sampling has to deal with problems of representativeness; the number of Internet users is limited, and is significantly reduced with advancing age. Researchers make an effort to overcome this weakness of Internet sampling by enlarging the population basis of prospective respondents, for example by means of sample triangulation, but the problem still remains. The extensive use of spamming and the resulting anger of Internet users makes a positive response to uninvited invitations a far from common incident. As noted above, offering rewards for full participation may be one answer to the problem, but even this has adverse effects on the quality of sampling.

Main points

- Sampling is the process of choosing the respondents and the units of the study in general.
- Sampling is a common practice and an indispensable research tool in social sciences.
- As the alternative to conducting a saturation survey, sampling offers many advantages.
- In quantitative research, sampling units are chosen prior to the commencement of the study, objectively and systematically; are easily identifiable and clearly defined; independent from each other; not interchangeable; and free of errors, bias and distortions.
- The two distinct types of sampling are probability and non-probability sampling.
- In probability sampling, all units have an equal, calculable and non-zero probability of being included in the sample.
- Non-probability sampling does not adhere to the rules of probability.
- Qualitative researchers employ non-probability sampling procedures such as theoretical sampling, accidental sampling, purposive sampling, quota sampling and snowball sampling.
- Usually, one sample is sufficient to conduct a study, but multi-sample studies are also common.
- Non-response is a serious research problem that investigators must deal with.

Where to from here?

Before you leave this chapter, visit the companion website for the fourth edition of *Social Research* at http://www.palgrave.com/sociology/sarantakos4e to review the main concepts introduced in this chapter and to test yourself on the major issues discussed.

Further reading

Etter, J. F. and Perneger, T. V. (1999) 'Snowball Sampling by Mail: Application to a Survey of Smokers in the General Population', *International Journal of Epidemiology,* 29(1): 43–48.

Fink, A. (1995) *How to Sample in Surveys*. Thousand Oaks, Calif.: Sage.

Gabor, M. R. (2007) 'Types of Non-probability Sampling Used in Marketing Research. "Snowball" Sampling.' *Management and Marketing,* 2(1).

Heckathorn, D. D. (1997) 'Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations.' *Social Problems* 44: 174–99.

Johnson, T. P. (2005) Snowball Sampling, Published Online: 15 Jul 2005.

Lavrakas, P. (1993) *Telephone Survey Methods: Sampling Selection and Supervision*. Beverley Hills: Sage.

Maisel, R. and Persell, C. H. (1996) *How Sampling Works*. Thousand Oaks, Calif.: Sage.

McPherson, M. (2001) 'Sampling Strategies for the Arts: A Hypernetwork Approach.' *Poetics,* 28(4): 291–306.

Sudman, S. and Blair, E. (1999) 'Sampling in the Twenty-first Century.' *Journal of the Academy of Marketing Science*, (27)2: 269–77.

Online resources

http://www.soc.washington.edu/users/brines/hardsampling.pdf