

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;

- how much variation there is in the population on the key characteristics being studied;
- the levels of measurement being used (nominal, ordinal, interval or ratio) and, hence, the types of analysis that can be applied; and
- the extent to which subgroups in the sample will be analysed.

A common misunderstanding is that a sample must be some fixed proportion of a population, such as 10 per cent.¹³ In fact, it is possible to study very large populations with relatively small samples; that is, smaller percentages. While large populations may need larger samples than smaller populations, the ratio of population size to an appropriate sample size is not constant. For example: for populations around 1,000 the ratio might be about 1:3 or 33 per cent (a sample of about 300); for around 10,000 the ratio may be about 1:10 or 10 per cent (1,000); for around 150,000 the ratio may be 1:100 or one per cent (1,500); and, for very large populations (say over 10 million), the ratio could be as low as 1:4,000 or a fraction of a per cent (2,500) (Neuman 2014: 270–1).

It is not uncommon for opinion pollsters to use samples of 1,000–2,000 with populations of many million. For such populations, increases beyond 1,000 produce only small gains in the accuracy with which generalizations can be made from sample to population, and beyond 2,000 the gains are very small (see Blaikie 2003: 176–7). However, the analysis undertaken in opinion polls is usually very simple. The factors mentioned above may require larger samples. It is in small populations that care must be taken in calculating the sample size, as small increases in size can produce big increases in accuracy, and vice versa. *It is the absolute size of a sample, not some ratio to the population size, that is important in determining the sample's ability to represent the population.*

Various formulae are available to calculate a suitable sample size. However, they are limited in their ability to take into account all the factors mentioned above. One approach in studies that work with sample data in percentages is to estimate the likely critical percentage as a basis for the calculation. For example, if we have an idea that the voting between two candidates at an election is going to be very close, then a poll prior to the election could select a sample to give the best estimate assuming each candidate will get about 50 per cent of the votes. Foddy (1988) has provided a formula for this.

$$\text{Sample size} = \frac{pqZ^2}{E^2}$$

where p is the expected percentage (say 50), q is p subtracted from 100 (in this case 50), Z is the t value for the chosen confidence level (say 95 per cent), and E is the maximum error desired in estimating the population parameter (say 5 per cent). In this example the sample would need to be 384, say 400.

Hence, two important factors enter into the determination of sample size: the desired accuracy – the tolerable sampling error – in the estimation of population characteristics; and the distribution of answers to a question, such as voting preferences. The higher the desired accuracy, the bigger the sample must be. For example, when a population is likely to be split about 50:50 in their answer to a critical question and the acceptable sampling error is 1 per cent,

a sample of 10,000 would be required (at the 95 per cent level of confidence). However, if the desired level of accuracy is 5 per cent with the same level of confidence, a sample of 400 would be sufficient. If the percentage split is anticipated to be 5:95, a sample of 1,900 would be required to achieve a 1 per cent level of accuracy and only 73 at a 5 per cent level, all at the same level of confidence. Hence, the anticipated distribution of answers to a question is also relevant. (See de Vaus 2002: 82 for a table that covers the range from 1 to 10 per cent accuracy.)

There is another way that the distribution of population characteristics can influence sample size. The characteristics of a relatively homogeneous population can be estimated with a much smaller sample than a heterogeneous population. Take age for example. If the population is all the same age, it is possible to estimate that age from a sample of one. However, a wide age distribution would require a substantial sample. Hence, the wider the distribution of a key population characteristic, the larger will be the sample required.

A third factor that influences sample size is the effect of the level of measurement and the associated method of analysis (see Blaikie 2003: 22–33). In general, the more precise or higher the level of measurement, the smaller the sample required, and vice versa. Interval and ratio measures require smaller samples than nominal measures. The reason for this is that nominal measures have to use cruder methods of analysis and, therefore, usually need larger samples to achieve satisfactory results. However, the distribution of population characteristics affects all levels of measurement, but they are more difficult to deal with when lower levels are used.

There is no simple rule of thumb for making the sample size decision. Before the study commences it is necessary to know what levels of measurement are going to be used and what methods of analysis can be applied to them (assuming they are both quantitative). Most good statistics textbooks will indicate what the minimum number is for using a particular statistical procedure, particularly for interval and ratio data. When results are presented in tabular form (usually cross-tabulations), a rule of thumb is that the sample size needs to be ten times the number of cells in the table. This rule is based on the requirement for chi-square analysis and the measures of association derived from it. The number of cells will be determined both by the number of categories on each variable, and by the need to meet chi-square requirements in terms of minimum expected frequencies. Other methods of analysis will have different implications for sample size. However, the complicating factor in all of this is the possibility that in any study a variety of levels of measurement will be used. The sample size will have to meet the requirements of the lowest level of measurement.

A fourth factor is partly related to the second and third. If it is intended that analysis is to be undertaken on a sub-sample, then the total sample must be big enough to allow for this. For example, if a study is conducted with a population of ethnic communities whose members migrated to a particular country before the age of eighteen, and if analysis is to be done on each ethnic group separately, then there must be a sufficient number of people from each group to do the analysis. Clearly, the size of the smallest community becomes important. The actual numbers required in this group will, again, depend on the levels of

measurement to be used and on the kind of analysis to be undertaken. If some of the variables are nominal, and three of these are to be used in a three-way cross-tabulation, the size of the group would need to be about ten times the number of cells in this table. For example, if country of birth (coded into three categories) is to be cross-tabulated with political party preference (three categories), and if the first variable were to be controlled by year of migration (coded into three time periods), then a table of twenty-seven cells would be produced requiring a sub-sample of 270. If the smallest ethnic community makes up 10 per cent of the population, then a total sample of 2,700 would be required. Of course, one way to reduce the total sample size would be to use a stratified sample by ethnic community, and different sampling ratios in each stratum, to make all sub-samples of 270. If there were five ethnic communities, the total sample would be exactly half (1,350), thus producing a considerable reduction in the cost of the study. However, if the variables to be analysed are interval or ratio, much smaller numbers can be used. One rule of thumb is to have a minimum of fifty in each subgroup, but, clearly, many things should be considered in making this decision.

Here are some important relationships between sample size, error and accuracy.

- As sample size increases, sampling error decreases and sample reliability increases.
- As population homogeneity decreases, sample error increases and sample reliability decreases.
- As sample error increases, sample reliability decreases, and vice versa.

'Sample size must take into account the degree of diversity in the population on key variables, the level of sampling error that is tolerated and the reliability required of the sample' (de Vaus 2002: 81). To reiterate, a small increase in the size of small samples can lead to a substantial increase in accuracy, but this is not the case for larger samples.

It will be clear from this discussion that a decision on sample size is rather complex. The best a researcher can do is to be aware of the effects of accuracy requirements, population characteristics, levels of measurement and the types of analysis to be used. The latter consideration reinforces the need to include in any research design decisions about how the data will be analysed. It is easy to think that this can be put off until later, but it cannot. Failure to make the decision is likely to lead to samples that are the wrong size, to data that cannot be sensibly analysed and, hence, to research questions that cannot be answered properly. In some studies, it is not possible to know in advance how the population is distributed on the characteristics being studied. Even rough estimates may be impossible to make. In this case, the researcher must be conservative and use a sample that will cope with the worst possible situation, which means making it larger!

Having said all this, one other major consideration enters into the equation. It is the practical issue of resources. The ideal sample needed to answer a set of research questions may be beyond the scope of the available resources. *Sample size decisions are always a compromise between the ideal and the practical, between the size needed to meet technical requirements and the size that can be achieved*

with the available resources. In the end, the researcher must be able to defend the decision as being appropriate for answering the research questions, given the particular conditions. If resources require that the sample size be reduced beyond minimum practical limits, then the design of the study would need to be radically changed, or the project postponed until sufficient resources are available.

It is always important to discover what conventions are used for your kind of research in your discipline or sub-discipline, in your university or research organization, and what the consumers of the research, including thesis examiners, find acceptable. These conventions do not always fit well with the technical requirements, but, in the end, may be politically more important.

Non-probability sample size

As it is not possible to estimate population parameters from the data acquired using a non-probability sample, the discussion in the previous section on confidence levels and acceptable errors in estimates is not relevant. If, however, quantitative analysis is to be undertaken, then sample size *will* be influenced by the requirements of the type of analysis to be undertaken.

When a research project involves the use of time-intensive, in-depth methods, particularly when directed towards theory development, the issue of sample size takes on a very different complexion. As we saw in the case of *theoretical sampling*, sampling decisions evolve along with the theory. It is not possible to determine in advance what the size should be. However, time and resource limitations will inevitably put some restrictions on it. In this kind of research, it may be more useful to think of selecting cases for intensive study, rather than getting distracted by sampling concerns that are irrelevant. It is to the discussion of case studies that we now turn.

8.7 Case Studies

Throughout the history of social research, *case studies* have been regarded in a variety of ways:

- as a type of research design;
- as involving the use of particular kinds of research methods, usually qualitative; and
- as being a method for selecting the source of data.

In the first view, case studies are included alongside surveys, experiments and ethnography/field research. As we saw in chapter 3, this way of classifying research designs is inappropriate. The second view goes back many decades to debates about the relative merits of survey research, with its statistical techniques, compared to participant observation and field research. 'Case study' was the collective term used for the latter. While the third view is probably the least common, it is the one that we want to emphasize here. 'Case study is not a methodological choice but a choice of what is to be studied' (Stake 2005: 443).