1 Sampling methods

This book is about the statistical aspects of sample surveys. However, it is mainly concerned with the statistics of samples that are not of the simplest kind and therefore an appreciation of the process of sample design and of the realities of sampling is an essential requirement for assessing the statistical approach necessary for any given survey. Before proceeding to the methods of dealing with data from samples which are not of the simplest and most straightforward kind, we must first understand why this additional complexity may be necessary, desirable or inevitable in the first place.

Accordingly, we devote this chapter to a brief review of the main methods of sampling and in particular to the thinking that underlies them. We do not intend to be comprehensive in scope or to be prescriptive. We shall cover only those issues affecting the statistical side of survey sampling, and not matters of data collection procedures, fieldwork administration or processing. For a fuller understanding there is a wide range of readily available texts on both the theory and the practice of sampling (see Appendix B for suggestions for further reading).

Survey sampling in the real world is frequently an imperfect process, and we must come to terms with its imperfections, not pretend that they do not exist. No sample used in commercial, scientific or academic studies will be a perfect representation of the whole from which it is drawn.

Correct procedures in the selection of samples will minimise the problems but they will not eliminate them. There is a temptation to think that, because a sample has been drawn according to the best advice and theory and using the best information available about the population or mass which the sample is supposed to represent, it can then be taken as a perfect (or even the best possible) microcosm of that population or mass. Similarly, the fact that a sample can be shown to be atypical in some respects does not necessarily invalidate the conclusions drawn from it, though it may prompt some re-examination of those conclusions.

In short, to use sample-based estimates properly we need an understanding of the factors at work in the sampling process, in theory and in practice, and at both the design and evaluation stages. We also need practical ways of assessing the effects of variations from the 'ideal' of the simple random sample.

Simple random sampling as the benchmark

The simple random sample lies at the heart of most teaching about statistics. It is the simplest form of sample to deal with mathematically; the associated formulae are generally compact and uncluttered. It has many elegant and desirable characteristics. It forms an excellent starting point for teaching the principles of statistical interpretation, reasoning and inference. Our concern is that, in the real world, the samples likely to be encountered by former students of statistics, and by researchers whose principal concerns and skills are not statistical, will not in practice be simple random samples. And many of these people will be ill prepared to deal with them except by treating them, incorrectly, as random.

The 'simple random sample' on which most statistics teaching is based implies that each and every unit in the population which that sample is to represent has an equal chance of selection, that each sample member is selected totally independently and that all selected members of the sample provide full information.

Many research projects call for a sample which is not 'simple' because the research has objectives for which 'simple' sampling is not the optimum. Except in artificial and controlled situations, most samples are not truly random because of the practical difficulties of selecting them and extracting the necessary information from them. This is particularly true of samples of the general human population. It is generally not possible to start with a comprehensive list or sampling frame of the population from which to make our random selection; it is often not possible to contact all the randomly selected members within the inevitable constraints of time and resources; and those contacted may not be willing or able to cooperate in the survey (an increasing problem). In a very large number of cases, therefore, we end up with a non-random sample even if we set out to select a random sample.

Even where a sample has been selected purely randomly, there is no guarantee that it will be representative, or comparable with another similarly randomly selected sample, in all respects, and some corrective action may be needed. Researchers who have set out to take a simple random sample may be tempted to feel that they have done their best and that they are absolved from responsibility for any shortcomings it may have. They have, however, an obligation to make themselves aware of any such imperfections and to deal with them appropriately.

The fact that a sample is not fully random, or is imperfect in its execution, and that the basic random sample statistical tools are not strictly applicable, should not be a cause for despair. There are ways of dealing with data from such samples in a professional and realistic manner and these are presented in the present volume. Nor should the fact that a sample has imperfections be in itself a reason for dismissing the results of a study using such a sample. The important things are to recognise what needs to be allowed for and to ensure that any interpretation and presentation of results makes a realistic allowance for the effects of non-randomness or unrepresentativeness.

This book is intended to complement rather than to replace the many books that deal with the standard and straightforward statistics of random samples. In this chapter, and in those that follow, we assume familiarity with the basic elements of probability theory and the simpler calculations involved with the evaluation of sample based estimates. (However, Appendix A contains a brief recapitulation of the fundamentals.)

1.1 Accuracy and precision

It is appropriate at the outset to stress the distinction between 'accuracy' and 'precision'. These words are sometimes used loosely or even regarded as interchangeable in the context of sample surveys. However, they are quite distinct concepts and researchers must be careful not to confuse them. In particular they should be aware which aspects of the conduct of a sample survey are likely to affect the accuracy of a survey's results and which will affect the precision – and which will affect both.

It must be remembered that sampling errors are not the only sources of uncertainty in survey based estimates. Results can be affected by other factors such as the wording or sequence of questions, respondent reaction to the subject matter (guilt, prestige), respondent fatigue, or even external influences such as the economic situation or the weather. These can have a greater and less predictable effect than sampling errors.

The 'accuracy' of an estimate denotes how well it corresponds to the true reality. Most surveys are carried out, naturally, to measure things where the true value is not known. The survey may provide the only estimate available. However, surveys also generally include measures where the true value for the population is known, even if these are only such basic facts as the age and sex of the sample members, and a comparison of such results with the known, true values may give some guide to the likely accuracy of other results.

Some measures for which the population value is known are used in the process of post-stratification and weighting, described later. Weighting is normally used to counter known biases, whether deliberate, as part of the survey design, or accidental as a result of imperfections in the sampling process. The variables used to determine the weighting are therefore forcibly brought into line with reality. Comparisons with population values for other variables should therefore generally be made after any weighting has been done. The variables used for post-stratification (explained in section 1.10) may also be compared with the true population values in their unweighted state, as a measure of the 'raw' sample quality, provided that due allowance is made for any differences arising purely from a deliberately disproportionate sample design.

Precision and reproducibility

Accuracy relates to closeness to reality and is largely dependent on the degree of bias involved. Precision on the other hand relates to the reproducibility of results. The precision of an estimate is largely dependent on the sample size, or, more strictly, on the *effective* sample size (a term that will be explained shortly). With a knowledge of the effective sample size it is possible to estimate how close to the present result a repeat of the survey would be likely to come, provided it was carried out in the same way and on the same scale. More specifically, we can predict how variable the estimates would be if the survey were to be repeated many times.

Precision can thus be more readily measured or estimated than accuracy. Merely repeating the sampling a number of times (or dividing the sample into matched replicates) will give some indication of the amount of variation to be expected in estimates and thus of the precision likely to apply to any individual estimate, even if we know nothing about the true population value. We can examine the size of

differences between an estimate from an individual sample and the mean of that estimate across all the samples. However, if we do not know what the true population value is from some other source, we have no way of estimating how far away from it our sample estimate is, whether from an individual sample or from a series of replicated samples.

Statistical theory gives us ways of estimating the likely variation between sample estimates using evidence derived from a single sample, but it provides no help with estimating how far any such sample estimate is from the true (but unknown) value. For that we may have no other recourse than to a judgemental identification of likely sources of measurement error and an equally judgemental assessment (where possible) of their likely direction and magnitude and of their relative importance, based solely on experience, logic and common sense.

As precision is predominantly a question of scale, it is possible for an estimate to be very precise, but also very wrong due to biases in the sample or in the method of data collection. Conversely it is also possible to be very close to the true result by chance with a small, poorly designed and ill executed survey – though not reliably so.

Sometimes precision may be needed more than accuracy. If the key measures in a survey are 'soft', for instance customer satisfaction (what does 'satisfied' actually mean?), then there is no way of assessing whether the result is 'accurate'. However, if a survey is be repeated over time the important thing is to ensure consistency of measurement so that whatever 'satisfaction' means we can be sure that we are measuring it the same way each time. Changes observed from one survey to another can be assessed to determine how likely they are to be indicative of genuine changes among the population rather than just part of the natural variation to be expected from repeated sampling. To do this we need to know what level of variation would be expected from survey to survey, and this should, ideally, be low relative to the magnitude of the changes observed. Here adequate sample sizes and above all consistency of methodology can be more relevant than strict purity of sample design.

At other times accuracy may be all-important. Predicting voting at an impending election, for instance, requires accuracy as well as consistency, because it really matters whether the final figure is 49% or 51%. Survey researchers are judged, rightly or wrongly but very publicly, on their ability to get such numbers right. A consistent difference of five percentage points may be acceptable in a customer satisfaction survey, but would be unacceptable in such a voting intention survey.

1.2 Design effect and sample size

The precision of any sample based estimate is determined by both the size of the sample and the way in which the sample was selected. The sample as used may provide a less or (occasionally) more precise estimate than a simple random sample of equivalent size would have done. In assessing the effect of sample design the variance of an estimate from the actual sample is compared with the variance of an estimate from a simple random sample of equal size.

Two terms that will be used extensively in this book are introduced at this point. The *effective sample size* is the size of a simple random sample which would yield an estimate with the same precision as the one we have. The *design effect* is the ratio of the variance[1] of the actual estimate to the variance of an estimate from a simple random sample of the same size. It may be more helpful to think of it in terms of the ratio of the actual sample size to the effective sample size.

A further term that may occasionally be encountered elsewhere is the *design factor*. This is the ratio of the *standard error* of an estimate to that which would have been yielded by a simple random sample of equal size. As such it is simply the square root of the design effect. The similarity of the name invites confusion without providing any great benefit and the term will not be further used in this book.

The design effect is a measure of the precision of an individual estimate. It is not a general measure of the precision of all estimates produced from a sample: every estimate produced by a sample has its own design effect and these can vary widely within a single survey data set. Nor should the design effect necessarily be viewed as a measure of the 'quality' of a sample. Samples that yield high design effects may do so because a complex design was required to achieve specific objectives. Nevertheless, where a simple sample design turns out to produce high design effects there is a case for asking why this might be.

The design effect is determined by estimating the variance of an individual estimate, and the estimation of variance is the subject of Chapter 3.

1.3 Defining the objectives

Designing the sample is one of the most important operations in conducting a survey. The principal decisions must be taken at a very early stage in the planning process. There is no single set of rules for designing samples and the design adopted in any instance will be the result of reconciling a number of often conflicting demands and constraints.

The design is (or should be) determined by the objectives of the survey. It is essential to have an agreed explicit statement of objectives in advance, and to ensure that these are realistic and appropriate. Project briefs for surveys are often written by people who do not understand the complexities of sampling. They may be insufficiently specific, lack essential information or be unrealistic in the implied expectations. This is not necessarily a criticism: the author of the brief may be someone very familiar with the subject matter and the issues and with the processes of policy-making or the management of a business, rather than a researcher. The main requirement, however, is to state why a survey is being done and what use is expected to be made of the results. It might be asked what business this is of the researcher; all the researcher needs is to

[1] The term 'variance' in this book is always used in its statistical sense of the mean squared difference between individual observations and their mean, and not in the sense used by accountants, of the (signed) difference between a reference value and an actual value (e.g. budget and expenditure or previous and current year).

be told what information is required from the survey. However, the researcher should be in a position to judge the extent to which the survey is likely to be able to meet the need of whoever initiates the project. A clear and agreed statement, in advance, of the background to the project, the objectives, the constraints and the means to be used to achieve the objectives will allow all parties to understand and think clearly about the practical issues. It may also bring to light any areas where different parties have different interpretations and hence different expectations, and it is helpful for all if these can be resolved before rather than after the survey is carried out.

Documenting the objectives

Even where the project is initiated by the researcher responsible for carrying it out, such a statement is essential to allow anyone examining the project in later years to see how and why the survey was conducted and, perhaps more pertinently, to assess the relevance and comparability of the results to other evidence. Re-analysis of existing data is often valuable and cannot be undertaken confidently without adequate documentation of the sampling process.

Where sample selection and fieldwork are to be contracted out it is essential that the potential contractors be given all possible relevant information. They need to know what they are expected to do and deliver and must be aware of anything known to the 'client' that is likely to affect the difficulty, cost or timing of the project. An experienced contractor may be able to contribute to the sample design process, if only in practical suggestions for making the process more cost efficient. Above all, a contractor must be certain that what is contracted for can be delivered on time and within budget.

It is therefore the researcher's task to ensure that before a survey is begun, the brief is set down as, or translated into, a clear plan which will allow the right decisions to be made about the sampling. Such a plan should set out:

- the objectives of the survey and the decisions likely to be affected;
- a definition of the population that the results of the research are intended to represent;
- the method of selecting the sample and, where appropriate, the sampling frame(s) to be used;
- the planned size of the sample;
- constraints of timing and budget, either or both of which may be negotiable to some extent;
- the nature of any important subgroups in the population who must be specifically represented;
- the subject matter of the survey;
- the time likely to be required by a respondent to complete the interview or questionnaire;
- the degree of accuracy or precision required in the principal measures;
- whether this survey is related to any other survey carried out in the past or simultaneously elsewhere, or planned for the future.

It may be thought that some of these matters have no bearing on sample design, let alone on the process of statistical evaluation of the results. But the subject matter of the

survey, for instance, may indicate or preclude some kinds of data collection method (thereby limiting the sampling options). Both the subject matter and the workload to be taken on by the respondent may influence the degree of cooperation obtained from those approached, also a sample design issue. Many of these matters need to be borne in mind when the precision of any estimate yielded by the survey has to be assessed. The design of the sample, particularly the method of sample selection, has a considerable bearing on the methods that can be used to calculate the variance of those estimates.

Reconciling priorities

Where there are multiple objectives, the priorities should also be determined. Sometimes the different objectives set out in an initial project brief may indicate different sampling approaches. If so, some objectives may have to be abandoned or relegated in priority or a compromise may be required. In reaching a compromise the relative priority of the different objectives should be indicated.

An essential factor, sometimes overlooked, is the general question of sample 'quality'. How good a sample do we need? It may sound heretical to suggest that we should ever accept, or even aim for, anything but the best, but in practice it is not always essential to strive for perfection. This is not a condonement of slipshod or incompetent work but a recognition that other factors may outweigh the requirement for top quality. The important thing is to match the quality of the sampling to other aspects of the task. If the principal information we are gathering is inherently imprecise (depending on the interpretation of question wording, for instance), then to strive for disproportionate accuracy in the sampling process may in some cases be less important than maximising the sample size.

There are times when a rough and ready sample is justified. In a pilot survey, the main concerns are often to establish whether the questionnaire 'works', whether respondents understand the questions and find them relevant and easy to answer, and to see what broad kinds of answers are given. Here getting a quality sample may be less relevant than ensuring that a good variety of people are interviewed. A pilot survey may, on the other hand, sometimes be used to see how well the proposed sampling method works in practice.

In other cases only a rough estimate of the incidence of some attribute may be needed; it may be sufficient to establish that the incidence of an attribute is so low in the population that it can be ignored or discounted, or sufficiently high that it cannot. Where a survey is intended to repeat a similar exercise done in the past so that the results can be compared and the degree of change estimated, then matching the previous method of sampling may be more important than doing it 'right', and the precision of comparisons will always be limited by the size of the initial survey. In such circumstances the method chosen to meet the objectives might be different from what would be chosen if there were no prior survey.

For some purposes, however, a correctly designed, carefully planned and rigorously executed sampling procedure is essential. Where crucial decisions of investment or policy hinge on the results, or where the results are to be subject to public or hostile scrutiny, we must have confidence in the results, and that confidence is founded on

knowing that the survey has been carried out to high standards. Sampling is largely under our control and we cannot afford to allow it to be the weakest link in the chain.

Probability and non-probability sampling

There is a fundamental division in sampling between 'probability' and 'non-probability' sampling methods. The division centres on whether or not it is possible to estimate the probability of a sample member's being selected. 'Probability' methods do not require that every member of the population have an *equal* probability of selection, but rather that the probability, or the relative probability, of selection of each member be known, either in advance or once the selection has been made. Whatever the actual process of selection, which can take many forms, knowing the probability of selection provides us with more confidence in the statistical treatment of the results. Statistical theory rests on known probabilities and a sound theory based approach is only possible for data sets where probabilities are known with at least reasonable certainty.

Non-probability methods should not be dismissed out of hand. Often there may be no practical alternative. If the objectives of a survey determine a method of data collection that makes a probability sample impossible, then a non-probability sample will have to do, and will often do very well. A farmer taking a sample of wheat to determine its moisture content does not laboriously select each grain individually or even ensure that all parts of the heap in the barn are sampled from. A scientist carrying out an experiment on sheep may allocate his batch of sheep randomly between treatment and control groups but he may have had little say in how that batch of sheep were selected from the total population of sheep. The difference (whether in money, resources or time) between the cost of a probability sample and a pragmatic non-probability alternative may be so great as to render the former unacceptable. A requirement to carry out all fieldwork within a critical and very narrow time-span may preclude full probability sampling methods. Where experience shows a method to be reliable, or where the overall method of selection can be shown (or even reasonably presumed) to have no or little bearing on the objectives, a non-probability sample may yield valuable and valid information.

What is certainly true is that more care and caution are needed in the interpretation of results derived from non-probability samples, and the precision of estimates derived from them is more difficult to estimate. However, the operative word is *more*. Results from even the best probability sample require care in interpretation and a constant awareness that samples are fallible.

There is a great danger in supposing that because a sample is a pure random sample it is therefore an accurate representation of the population it represents. The researcher may have done everything possible to ensure that no non-random influences have occurred, but should not on that account be complacent about the quality of the results. The most carefully drawn sample can still turn out to be noticeably biased through sheer bad luck. Two identically drawn random samples will invariably yield different results, so at least one of them is 'wrong'. There is also the question of non-sampling errors, which this book does not attempt to cover, and which can apply to probability and non-probability samples alike.

1.4 Defining the constraints

Sample surveys are practical exercises and have to be carried out in the real world. The theory of sampling and sample design has to be tempered with the practicalities of locating and extracting information from the prospective sample members.

A sampling plan put together without an understanding of the limitations of the materials available, and most particularly of the real operating conditions under which fieldworkers collect the data, is likely to lead to unrealistic expectations on the part of those commissioning the survey. It is also likely to lead to failure on the part of those conducting the survey to deliver in accordance with the (unrealistic) specifications. At worst it may result in hasty corrective action being taken mid-way through a research project, as problems become apparent, to modify the methods being used.

It is therefore in everyone's interest to ensure that the constraints within which a survey will be carried out, over and above those implied by the project plan, are understood in advance. The sampling plan should therefore contain statements of the assumptions made in arriving at the recommended design, and the constraints within which the plan has to be executed. Such constraints may include:

- limitations imposed on the choice of data collection methodology by the objectives or subject matter or by logistic restrictions;
- the availability of sampling frames, the extent to which these are comprehensive, up-to-date and reliable, and the level of detail they offer;
- where relevant, the information available against which the results delivered by the sample can be assessed;
- any legal, ethical, cultural or similar considerations affecting or restricting the conduct of the survey;
- the need to conform to any precedent, or to follow the design or methodology of another piece of work to allow comparisons to be drawn.

Cost and quality: the trade-off

Some apparent constraints may be negotiable (e.g. time and budget) or otherwise less firm than they may appear, where it can be shown that relaxing them would lead to a 'better value' end result. It is often advisable to consider, and possibly to table for discussion, more than one plan before a decision is made about the sample design. It is certainly essential to know how variations in the elements of the design or in the assumptions affect the final costs as well as the quality of the survey. For instance a response rate 5% below what was planned will probably have an adverse effect on both cost and quality, while requiring one additional interview per cluster (see section 1.12) may save money but could reduce the precision of estimates.

The issue of the trade-off between cost and quality is too complex to be covered systematically in this book. Most kinds of survey errors related to sampling can be reduced, at a cost. For a wide-ranging review of the questions of errors and costs see Groves [24].

But assessing the extent of such reduction is not a simple matter. The cost and the likely degree of reduction have to be weighed carefully, both against each other and

against the objectives of the survey and the kinds of measures made by it. Striking a balance is not easy and researchers with different priorities or interests may differ considerably in their judgements of what should be sacrificed to what. It is very easy to criticise a sample (usually someone else's) for being inadequate in design quality or insufficiently rigorously executed, or even for being 'over-engineered'. The poor researchers responsible for making these decisions are at the mercy of both the ivory tower academic and the accountant, but should at least be satisfied in their own minds that their decisions with respect to sampling are guided by the principle of 'fitness for purpose', however they see that purpose.

1.5 Defining the population

At an early stage in the planning of the project a decision must be taken about the population to be studied. The *population* (sometimes referred to as the *universe*) is the totality that our sample is to represent. This is often made up simply of people, but it could also be a population of households, businesses, institutions, ships, shops, sheep or cars, or of events such as journeys or admissions to hospital. It is important to define the target population precisely. Is it just 'all people' or 'all adults' (if so, what defines an adult?) or some more specific group, such as 'persons aged 18–64 living in private households'. The target may of course be a very specialised population, such as 'all practising dentists' or 'all dental practices'.

Existing information about the size and nature of the defined survey population should be reviewed, as this may provide guidance for sample design. Are its members disproportionately concentrated in certain regions, or likely to be found in certain neighbourhoods? Can they be readily distinguished from other members of the population? Is there, in extreme cases, an available list of its members?

The 'survey unit' should also be defined. Are we surveying individuals about themselves, for instance, or are we surveying households or businesses but obtaining the information from one person in each?

In reality it may not be possible to ensure that the target population can be fully covered in a survey. There are groups in the general population which in any given survey it may be impractical to reach adequately or at all. These may include, depending on the circumstances:

- people in remote or sparsely populated areas (for logistical and cost reasons);
- people not in private households (people in hospitals, prisons, convents, boarding schools, at sea, etc.);
- people who are regularly away from home for long periods;
- people without telephones;
- people too mentally or physically incapacitated to undertake an interview;
- people with insufficient command of the language in which the survey is conducted or, where appropriate, inadequate literacy skills.

There is also the increasingly large group in the general population of people who simply *refuse* to cooperate in a survey. Unlike many of the other groups, this one may