1

ALL THE WORLD'S A SAMPLE

Introduction

Reading almost any archaeological document, from a research proposal to a fieldwork report, it will not be long before one comes across the word 'sample' in one context or another. A project, whether at regional or site level, will be based on a *sampling strategy*, a research design may specify that features are to be *sampled*, and *samples* of various types will be taken for the delight or otherwise of specialists who wish, or who can be persuaded, to look at them. In fact, almost all archaeology involves sampling; indeed, one could say that there is a sense in which much of archaeology *is* sampling, echoing David Clarke's remark that 'Archaeology . . . is the discipline with the theory and practice for the recovery of unobservable hominid behaviour patterns from indirect traces in bad samples'. (Clarke 1973, 17).

The word sample is all-pervasive, but one soon comes to realise that it does not have the same meaning each time it occurs. This is not surprising, as just one of the six definitions given by the *Oxford English Dictionary* is broad enough to encompass a wide range of meaning: 'a relatively small quantity of material, or an individual object, from which the quality of the mass, group or species, etc. which it represents may be inferred'. At one extreme, one may encounter a 'multi-stage probabilistic sampling strategy', while at the other, one may encounter the casual use of the word sample to refer to collections of muddy objects in plastic bags. Cynically, one might note the existence of piles of 'samples' in dark corners of archaeological stores, whose main role seems to be to get in the way for several years, and then to be thrown away. The word sample has become, through usage, so diffuse that we need to bring it back into focus before we can start to understand what it means, where it comes from, and what implications it carries for the way in which we actually do archaeology.

To try to achieve this focus, we need to split apart the various meanings that have accreted on to the word sample, and look at each separately. As a first step, we can divide samples into:

(a) *Unintentional samples.* Here the sampling has been done, so to speak, before the archaeologist arrives on the scene. We know full well that the material we have painstakingly recovered is not the whole of what was lost or discarded in the course of activities undertaken at its location. Sometimes its condition makes this obvious – for example sherds are clearly only a fraction of the pots that they represent – but at other times its condition may lure us into thinking that we have a 'total sample' (e.g. of coins). But survival does not guarantee recovery; both size and colour, among other factors, may affect the chance of detection, thus imposing a curious form of sampling based on an archaeologist's ability (combined with weather, deadlines, etc.).

(b) *Informal samples*. Next we come to samples whose selection is, to some extent at least, in the deliberate choice of the archaeologist. The choice may be based on archaeological criteria or on those of time, cost and convenience, or on a combination of them. There is a spectrum of intentionality within this group: at one end we might put *purposive samples*, for example carefully targeted excavation units based on topographical features or geophysical survey. At the other end, we might put *haphazard* or *grab samples*, for example the hasty gathering-up of a few objects found on the surface of a potential site. Somewhere between the two might come the idea of a *typical sample*, selected by the archaeologist to represent a collection of objects, though the achievement of typicality is far more difficult than might be expected. What all such samples lack is the potential for generalisation from them: the ability to move from the description of a sample to a reliable statement about some wider entity, usually known as a *population*. Such power is provided by:

(c) *Formal samples*. These are samples selected from well-defined populations according to rigorous statistical procedures. If taken properly, they enable us to make valid statements about the relevant populations, such as estimates of certain parameters (e.g. density of sites in a region, size distribution of inclusions in a ceramic fabric). Equally importantly, they can provide likely margins of error for such estimates, which can not only tell us how 'useful' they are likely to be, but can also help us to determine how large a sample has to be to tell us what we want to know (at least, with a chosen level of confidence). The price of this power is exposure to, or even immersion in, statistical sampling theory, which may seem rather daunting to many archaeologists, but which in many archaeological situations is unavoidable.

It might seem that I am here making the equation: formal = good, informal = bad, but that would be a gross over-simplification. Certainly, there are situations in which this is true: the subjective choice of sampling units, whether at a regional scale or in a museum store, can easily lead to bias (p. 23) and an inability to quote reliable margins of error for one's results. On the other hand, formal methods are to some extent a rational response to a state of ignorance, and the more we know about a situation, the less necessary they may be (Redman *et al.* 1979). For example, to ignore evidence from aerial photographs or geophysical surveys in designing a purely random sample of (for example) test-pits would be wasteful and unproductive, and a more targeted approach would be likely to give more useful results. But there is a contrasting danger, that of the self-fulfilling prophecy – looking only at areas where one expects to

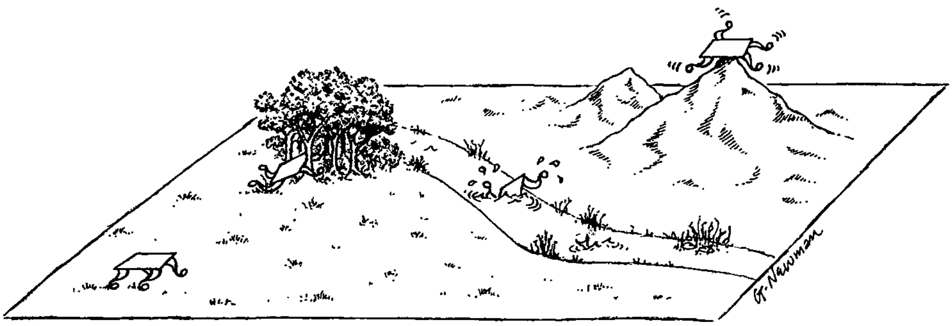

Fig. 1.1 Invasive sampling or the march of the quadrats.

find 'something' may merely confirm what one already knows – and some (lesser) attention should be paid to apparently 'blank' areas. So there needs to be a balance between statistical rigour and the use of what statisticians call *prior information*, in order to make best use of resources and to achieve reliable outcomes. One approach to statistics, known as *Bayesian statistics* (p. 16), seeks to make explicit use of such information, but it requires a higher level of expertise, particularly in modelling, and is only just beginning to make an impact on survey sampling. But even if such advanced theory is not accessible, there is a need to break away from the mental image that equates sampling with an army of small squares marching willy-nilly across the landscape, oblivious of rivers, mountains and any other natural features (Fig. 1.1). This caricature is born of very incomplete understanding of sampling theory, which results in archaeologists using it as a self-imposed straitjacket, against which they rebel. Sampling theory is a good deal more flexible than is sometimes supposed, as we shall see in the following chapters.

First, though, we need to continue to break down the definition of sampling, this time according to the *scale* at which it takes place. In the course of fieldwork, we may be called on to sample from a region, a site, or a feature on a site. In the laboratory, we may need to sample from an assemblage of excavated material, or from within the material of an object itself, such as a pottery sherd or a metal artefact. Even when laid to rest in a store, artefacts may not be free from our attentions, as we may decide to sample them to check that they are not decaying at an unacceptable rate. Activities at these different levels tend to be undertaken by different groups of people, who have built up their own professional expertise and case law. This has tended to obscure the underlying unity, that at whatever the scale it is all *sampling*, and shares common problems and possibly solutions. The unity is provided by statistical sampling theory, which at its core is context free, but which nevertheless seeks to adapt itself to the needs and assumptions of particular situations. A secure grounding in theory can enable one to see where an advance made at one scale can be successfully transferred to another (and, just as importantly, where it cannot). In a discipline which is well known for its penchant for borrowing techniques from other disciplines (p. 171), it is perhaps surprising that there has been so little cross-fertilisation between the sub-disciplines that make up archaeology. Perhaps the reason lies in archaeologists' ambivalent attitudes towards sampling, which could provide the link.

Attitudes and history

To try to untangle the complex web of feelings and opinions that archaeologists have about sampling, I have created the following series of thumb-nail sketches. Obviously they contain an element of caricature, and archaeologists may well recognise an element of themselves in more than one of them. Nevertheless, they can form a useful basis for exploring the shifts in opinion and practice that have occurred over the years.

(a) *Mother's milk*. Sampling is something that all archaeologists do all the time, and as such they instinctively know all they need to know about it. This view is often tacit rather than articulated; when it is expressed, it may be in the form of 'professional judgement'. Found mainly in commercial archaeological units.

(b) *Black art*. Sampling could be the answer to all our problems, but someone somewhere is withholding a vital piece of information. A magic percentage of our site, region, etc., will tell us all we need to know, if only we knew what that percentage is. Second to (a) amongst field archaeologists.

(c) *Alien imposition*. Sampling theory is an attempt by statisticians (and like-minded people) to impose their values and methods on an unwilling discipline. Their methodology is inappropriate because archaeology is unique and has nothing to gain from such outsiders. Found amongst those who know enough statistics to recognise the folly of some archaeological sampling schemes, or amongst those who are simply scared of it.

(d) *Regrettable necessity*. 'Had we but world enough, and time' we would do everything – total survey, total excavation, total record. Unfortunately, we don't, so we have to compromise with the constraints imposed by real life. Sampling is a sell-out, but until we get proper funding, etc., we have to live with it.

(e) *Passport to respectability*. All respectable scientists sample, so if we are to hold up our heads in the scientific community, then we must do so too. The more explicit the theory the better, and admire my research design. Main habitat is research grant applications.

(f) *Escape from tedium*. Much archaeological activity consists of the tedious compiling of mountains of low-grade data. Surely we don't need it *all*? Why not base our analyses and interpretations on just a sample of it, and spare ourselves the effort of measuring another (large number) of (chosen artefact type)? Particularly common amongst finds specialists.

(g) *Framework for research.* We need to ask 'how much work (fieldwork, finds study, etc.) do we have to do to find out what we want to know?' Ask questions, design research to answer them, sample, analyse and interpret, then move on. Target resources carefully.

The history of sampling in archaeology will be examined thematically in chapters 4 to 8, but here we look at it in terms of the attitude behind the practices. Initially, sampling was seen as an intuitive exercise (a) – if, for example, there were more sherds on the surface of a site than you could reasonably collect, then you took what seemed to you to be a 'representative' sample (p. 112). More formal methods started to come in in the 1940s, after the wider dissemination of statistical sampling theory that followed the theoretical advances of the 1930s (p. 15) and their application to quality control in response to the industrial pressures of World War 2. They were used first on dense concentrations of data in small spatial locations (e.g. shell mounds, pp. 145–7), seeming to be motivated by attitude (f).

A turning point in many ways was signalled by Binford (1964), although technically precedence should be given to Vescelius (1960) (p. 68). Binford advocated three changes, one archaeological and two statistical: the promotion of the region to the position of primary unit of archaeological research, the explicit *design* of archaeological research programmes, and the explicit use of sampling theory as the way of linking the two. Although his practical example was rather contrived, this paper had a tremendous impact and did in fact achieve its apparent aims. Certainly, in fewer than ten years, formal sampling had reached such a level that a major symposium could be held on the subject (in San Francisco 1973, see Mueller 1975b), and a series of textbooks and reviews soon followed (see pp. 69–70). It is worth noting that the majority of contributions at San Francisco were on regional sampling, with a minority on site sampling and only one on sampling at a smaller scale. In contrast, at the British counterpart (in Southampton 1977, see Cherry *et al.* 1978), there was a much more balanced representation of the various scales.

Inevitably, there was opposition of one kind or another. At the feature level, the widespread introduction of sieving methods in the 1960s and 1970s led to a growing realisation that sampling was essential to keep the resulting workload down to a manageable volume. At first, this seems to have been done reluctantly, in the spirit of (d) (e.g. Payne 1972a), but later more enthusiastically (e.g. Veen and Fieller 1982; Levitan 1983). At the site level, moves towards more formal sampling were paralleled by realisations that, to answer some questions, larger rather than smaller areas had to be excavated (Biddle and

Kjølbye-Biddle 1969). Although this was initially a reaction against the 'Wheeler system' of excavation, in which a surprisingly high proportion of a site was left unexcavated as baulks, it led to expressions of the view that total excavation was the only 'proper' excavation (e.g. Barker 1977; see Cherry *et al.* 1978, 151). The over-rigid use of sampling designs at regional level was quite rightly criticised (e.g. Hole 1980; Wobst 1983), though at times the concern seems to have been with the idea of sampling rather than with the way in which it was being employed (c).

By the 1980s in the USA and the 1990s in the UK, sampling had become firmly entrenched in the methodology of the 'contract' wing of archaeology (pp. 69, 74), to the extent that one could almost say that the wheel had turned full circle and returned to (a), though at a higher level of self-awareness, and with an element of (b).

Where this book stands

This book is written from a background of an academic training in mathematics and statistics, favouring attitude (g), and an archaeological career spent mainly in ceramics, a material which leads one towards attitude (f). Attitude (e) is perhaps less important – archaeologists must find a methodology that is appropriate for them, and should not be always looking over their shoulder and worrying about what others might say. From this perspective, sampling is not an option but an imperative that should be embraced willingly, not reluctantly. It is not a 'second-best' strategy forced on us by lack of resources, but a responsible use of whatever resources may be available to us, whether small or large.

The root of the issue is that archaeological fieldwork has the potential to generate enormous, indeed overwhelming, quantities of data. These quantities need to be limited, for both logistical and intellectual reasons. On the logistical front, we run immediately into problems of resources – money, time and scarce specialist expertise. Many fieldwork projects are heavily constrained by the need to obtain the required information (often for external purposes, such as local government planning control) from a very restricted budget in a competitive environment. It would be wrong, from many points of view, to retrieve more material than we can handle, and simply trying to increase our resources does not help, since it will (at least in the short term) only decrease resources elsewhere. The 'free lunch' of volunteer labour is not the answer, since it raises the questions of whether such workers would be better employed in doing something else (either within archaeology or outside it), and of whether we are exploiting them. Finally, all this material must go somewhere; museums in the UK are in a storage crisis brought about in part by the boom in archaeological fieldwork since the 1970s. These arguments involve the whole of archaeology, from 'cradle to grave', one might say; the point is that sampling is unavoidable, and that informed awareness of formal methods gives us the best chance of obtaining the results that we require, from the resources available.

The intellectual arguments are at least as compelling. There is no point in accumulating data, either for their own sake ('stamp collecting') or in the fond hope that they will answer someone else's questions one day. When or if a new 'wonder' technique does arise, the chances are that it will not be our laboriously catalogued data or stored material that will answer it. For example, when thermoluminescence was developed as a method of dating pottery sherds, it could not be used retrospectively because information about the nature of their surrounding matrix was rarely if ever available (Aitken 1990, 154). Next, data are subject to a law of diminishing returns. If we double the quantity of data, we do not double the information that (even in principle) we could obtain from them (p. 210). Sooner or later, as the volume of data increases, there comes a point at which further data provide virtually no extra information (e.g. p. 157). Another point is that, within fixed resources, sampling allows us to pay more attention to our material, and so improve the quality of our data. A relatively small quantity of good-quality data is often of more use than large amounts of low-quality data, as many opinion poll specialists would testify. Finally, too many data can actually obscure rather than illuminate any patterns that may be present. They can do so either by sheer weight of numbers, so that we 'can't see the wood for the trees', or because the irrelevance of some of the data can obscure patterns in the more relevant data (p. 181).

How, then, can we limit the quantity of data that we retrieve and accumulate? In a very simple model, we can see data as measurements or other observations made of *variables* on *objects* (using *object* to refer to any archaeological entity, and *variable* to refer to any characteristic that may vary from one object to another: see below for a more detailed discussion). Under this model, we can reduce the overall quantity of data either by reducing the number of objects (i.e. by sampling) or by reducing the number of variables (which is here called *selection*). Selection is important because it is often the *irrelevant* variables that obscure patterns, while *redundant* ones (i.e. ones that contribute no extra information) just waste resources. But the issue of selection may make archaeologists feel uncomfortable or insecure. Fortunately, it can to some extent be managed by the use of more appropriate data structures or models. For example, in the 1970s British ceramic specialists, under the influence of Peacock (see p. 178) and feeling their way in relatively new territory, recorded many details (e.g. nature, size distribution, frequency and shape of inclusions) about sherds individually, in the belief that it could all be sorted out on the computer (urged on by the siren-like encouragement of data-processing specialists with little connection with archaeology). By and large, it couldn't be, and much effort was wasted. The correct place to sort out ceramic fabrics is at the work-bench: fabrics can be defined (using ranges of values rather than point values of variables) and sherds can be seen and recorded as examples of particular fabrics. This structure of the data mirrors that of the *relational database*, which was coming into general use at about the same time, and which advocated the superiority of several small tables of data over an all-embracing individual one (a lesson that has not been entirely taken on board by all archaeologists). However, improved structuring of the data does not remove the archaeologist's responsibility to select variables sensibly in the first place.

Given that sampling is essential, we need to ensure that it is carried out in a way that enables us to make best use of the data that it provides. The requirement is that a statement that we make, based on data from a sample, should in some measure be true of the corresponding population, and that we should have some idea of how accurate it is likely to be. This is usually expressed by saying that the sample should be *representative* of the population. In terms of strict sampling theory, this is achieved by ensuring that the sample is *random*, i.e. that each element of the population has a certain, known, chance of being chosen to be part of the sample. This requirement, and the variations in sample design that are possible in different circumstances, are described in detail in chapter 2. The problems that arise in reconciling formal statistical requirements with archaeological reality form the 'meat' of this book in chapters 4 to 8. As we have seen in the introduction to this chapter (p. 2), formal sampling does not seem to be the appropriate answer for all archaeological questions. For example, if one wished to date a ditched enclosure, one or more sections cut across the bank and ditch would be a preferably strategy to a random set of trenches across the whole site. If, however, one wished to estimate the density of activity within the site (measured perhaps in terms of the density of features or artefacts), then the latter *would* be appropriate. The difference is that, in the latter case, we wish (in statistical language) to estimate a parameter of the site as a whole, while in the former the estimate we want is based on material that, experience tells us, is more likely to be located in certain parts of the site. The important point is not so much whether a sample is formal or informal, but whether it is *designed* or *undesigned*. Designed samples include formal statistical (random) samples, which are appropriate for many questions but not for all. Obviously, in this book we shall concentrate, but not exclusively, on those situations in which formal sampling methods have the most to offer.

All this discussion of data raises the question of what data actually *are*? The line taken here is that they are characteristics of objects in the real world, which archaeologists can choose either to observe and record or not to observe and record. The role of the archaeologist, and of archaeological theory, is to select relevant variables according to the question in hand, but not actually to create them. Data, of course, are far from perfect. The values that we record depend on a wide range of extraneous factors, over and above the 'true' value of *that* variable on *that* object. Personal and environmental factors can influence the perception of variables such as colour, and even apparently 'objective' variables, such as length or diameter, can be subject to a surprising level of variation, plus the occasional gross recording error, due perhaps to transposed digits or an omitted decimal point. Gross errors can usually be detected relatively easily, because they stand out; more difficult are small systematic errors that may reflect personal biases. For example, I once analysed differences between rim diameters of pots of the same form found on different parts of a production site, only to find that they had been caused by psychological differences between two recorders, one of whom rounded doubtful diameter measurements down to the nearest inch, while the other rounded them up. The allocation of tasks to different workers, to prevent this sort of problem, requires far more attention than it is usually given (but see Daniels 1978).

The use of instrumentation, even expensive instrumentation, does not remove the possibility of error, but brings its own problems. Equipment used in many sorts of elemental analysis may suffer from 'drift', and have to be calibrated from time to time, while rare elements are subject to a 'threshold' effect: values smaller than a certain amount are recorded as zero. If different elements have different thresholds, or (in a comparative study) different instruments have different thresholds, then analysis of the data will at best be very difficult.

In all cases, then, it is important to maintain a critical attitude to one's data, remembering the old adage 'if it looks wrong, it probably is wrong', or even Twyman's Law that 'any figure that looks interesting or different is probably wrong' (Ehrenberg 1975, 10). This is especially true once data have been stored on a computer: the process of data entry itself can create further errors, but data seen on the screen or on a print-out have an annoying way of looking 'right' simply because they are 'in print'.

The quality of data is a general statistical issue, although it also includes aspects that are specific to each discipline. Much work has been done, and continues to be done, on ways to prevent, detect or minimise errors in data, and, although it falls outside the scope of this book, archaeologists are advised to be aware of it.

All that remains in this section is to bring together this view of data with the role of sampling in archaeology, to present a coherent model for the process of archaeological research as a whole. Fig. 1.2 shows just such a model, which is equally applicable to any other data-based discipline. It is quite complex, and needs to be 'unpacked' carefully. The model first makes a distinction between the world of theory (what goes on inside your head) and the real world (what goes on outside your head). Theory is characterised by *hypotheses* (or, more generally, ideas or opinions), and the real world by *data*. It claims that the two interact not directly, but through an intermediary, the *model*. As a simple everyday example, consider the problem of leaving home in time to catch a

Cambridge University Press
978-0-521-56666-7 - Sampling in Archaeology
Clive Orton
Excerpt
More information

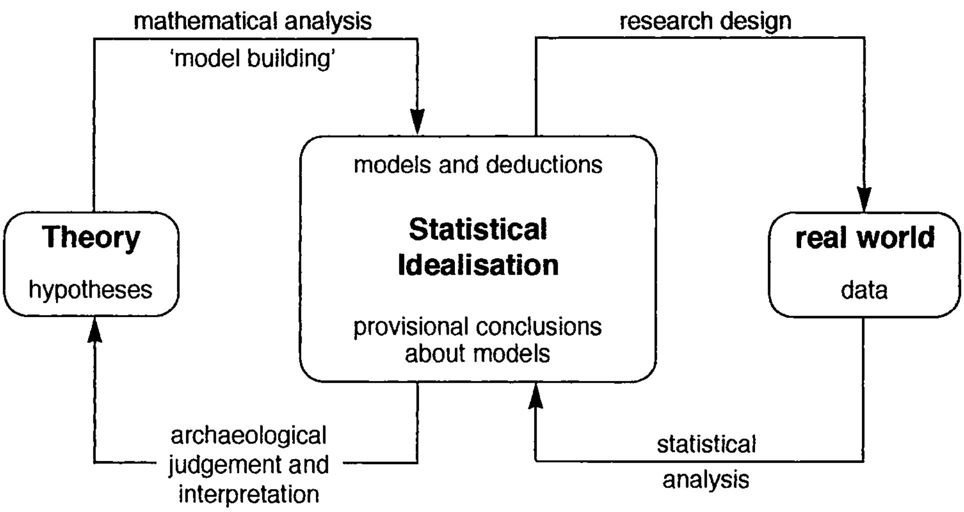

Fig. 1.2 The research cycle.

train. The train (representing the real world), will arrive at the station at a definite but (to me) unknown time. My theoretical task is to judge when I need to leave home. The actual position of the train is of no use to me, because I have no way of knowing it. But what I do have is a model, in the form of the railway timetable plus my experience of how long it takes to walk to the station. Initially, this model provides me with my best guess of when to leave home. Later, it may become modified by experience; if I discover that the train is always at least five minutes late, for example, I may dare to leave home five minutes later. This simple example makes some useful points about models; first, that they are *simplified representations of reality*, and second, that they can be modified in the light of experience.

Having established the main building-blocks of our model, we can now start to see how they relate to each other. The first point to make is that there is no 'correct' starting point: we start from where we are. If we start by thinking about a problem, we are in the world of theory; if we start with a collection of objects, we are in the world of data. For purposes of illustration, we shall start from theory. The first step is to create a model of the situation that we wish to investigate. This may consist of defining the population that we wish to explore, and the variables that we wish to observe. The key questions may well be which population and which variables are relevant to the question that we wish to answer. This step is more difficult than is often acknowledged and may be glossed over. The next step is the well-known one of research design: deciding on the size and design of our sample, and the practical issues of its implementation and analysis. We can then go out to the real world and collect our data. The third step, statistical analysis, should be a formality, since we should have decided how we are going to analyse our data *before* we collect