17 Sampling in Geography

Stephen Rice

Synopsis

Geographers recognize the value of both extensive statistical sampling and intensive 'case-study' sampling for exploring an uncertain world. A benefit of extensive sampling is that a set of techniques known as inferential statistics can be applied to make probabilistic statements about the population from which the sample is drawn. Sampling is therefore a powerful tool, but geographical research frequently engages with heterogeneous phenomena that require careful sampling in order to maximize the accuracy of conclusions. Thoughtful design of the sampling programme is therefore crucial and is driven by both the research aims and available resources. The chapter is organized into the following sections:

- Introduction
- Samples and case studies
- What makes a good sample?
- Designing a sampling programme
- Statistical inference
- Sample size
- Conclusion

INTRODUCTION

Sampling is the acquisition of information about a relatively small part of a larger group or population, usually with the aim of making inferential generalizations about the larger group. Sampling is necessary because it is often not possible, practicable or desirable to obtain information from an entire population. For example, it is essentially impossible to measure all of the sand grains on a beach to ascertain their average size; impracticable, in the course of a normal day's work, to question every person on the beach to determine the variety of their views about personal use of public spaces; and undesirable, never mind unethical, to stress the fish community of a seashore rock-pool by catching and examining all of its members. Moreover, in quantitative research, a set of procedures known as inferential statistics can be applied to sample data in order to make generalizations, validated by probability statements, about the entire population from which the sample was drawn. Thus, it is not *necessary* to interrogate a whole population to make useful generalizations about it.

In one form or another, sampling is the basis of almost all empirical research in both physical and human geography and is widely relied upon. However, such a powerful methodological tool comes with a set of health warnings: samples are only as valuable as they are representative of the larger population; at best, bad sampling leads to imprecision; at worst, bad sampling yields incorrect or prejudiced results. In the extreme, lack of sampling rigour makes inferences meaningless: 'The tendency of the casual mind is to pick out or stumble upon a sample which supports or defies its prejudices, and then to make it the representative of a whole class' (Walter Lipman, journalist, 1929). This remark may be directed at the dangers of using isolated examples to make unfounded inferences in everyday life rather than in academic research, but it nicely highlights the important association of weak, non-rigorous sampling with bias and inaccuracy.

This association, the difficulties that are frequently encountered in order to collect representative samples and the impenetrable nature of 'stats' for some students breed a scepticism about sampling and statistical inference that can run deep. It is fairly common to hear the dismissive statement 'Well, you can prove anything with statistics.' Benjamin Disraeli's remark that 'There are three types of lies: lies, damned lies and statistics' has been a convenient aphorism of opposition politicians concerned with the scruples of their counterparts for over a century. This scepticism is, however, unfortunate because geographers often seek to understand spatially diverse, highly complex phenomena that can be efficiently accessed by sampling and usefully examined using statistical inference.

While questions about motivation and affiliation (political, epistemological, etc.) have a place in the critical assessment of all research, for students of method the value of statistical inferences must be judged simply against the quality of the sample data and the quality of the analysis applied to them. These issues are the focus of this chapter. It reviews the use of sampling by geographers; considers the criteria by which samples should be judged; discusses the design and implementation of sampling schemes; introduces some of the basic elements of statistical inference; and suggest some methods for defining sample size. The overall message is that sampling is a powerful tool that most geographers need, but if research methodology aims to be as impartial and free of error as possible, then sampling must be done thoughtfully and rigorously.

SAMPLES AND CASE STUDIES

Not all geographers seek to make quantitatively robust inferences about large populations or develop theories and models of universal validity. Nevertheless, a principal aim of most, if not all, geographical research is to make useful generalizations – that is, to seek out and explain patterns, relations and fluxes that might help model, predict, retrodict, or otherwise understand better, the human and physical worlds around us. Thus, some geographers may restrict their attention to small areas, short periods of time, small groups or even to individual places or people, but the underlying approach remains nomothetic – it focuses on identifying the general rather than the unique, and in turn most geographical research involves some form of sampling.

Geographical systems are complex and affected by historical and geographical contingencies, indeterminism or singularity (Schumm, 1991). That is, while there are general similarities between objects, there is also inexplicable (or at least, as yet unexplained) variation in any population, so that each item is a little different from the others. This means that while we may be able to develop generalizations that are valid for a whole population, estimating the behaviour or character of any single individual is difficult and prone to error. This is as relevant to predicting river discharge as it is to commenting on the global reach of large corporations or people's views on street architecture. Thus, the geographical world is an uncertain world.

Geographers have adopted a variety of strategies to search for general understanding while recognizing this uncertainty. A useful distinction can be made between approaches that intensively examine a small number of examples and those that sample extensively (Harvey, 1969; Richards, 1996). In the extreme, geographers have made substantial use of single case studies (samples of one) to learn about both physical and human phenomena. While case studies provide detailed information, a fundamental criticism of the approach is that the generality of the case is unknown. That is, there is no formal basis for substantiating inferences made about a population on the basis of a single sample, such that extrapolating the findings of the case to the population remains merely a matter of intuitive judgement on the part of the investigator. This is problematic, because most geographers would (and should) be sceptical of such subjectivity. In contrast, the collection of large, extensive samples provides the opportunity to utilize statistical methods which, at least when sampling is appropriately conducted, offer a means of assigning objectively derived conditional statements to population inferences.

While recognizing the limitations of case studies and the benefits of statistical inference it is important not to denigrate the value of intensive investigations, not least because there is no single, simple model that defines how geographical knowledge can or should be obtained and ratified (see Chapters 16, 18 and 19). Case studies may not provide a basis for making wide-ranging inferences about a population but inferences based on case studies are not necessarily false or unreasonable. Rather, they stand to be substantiated and the detailed information gathered in a case study may reveal general structures or relations that can be used to generate or modify models or hypotheses (Harvey, 1969). Similarly, case studies may present unique opportunities for understanding the mechanisms that underlie empirical observations. In geomorphology, Richards (1996) argues that as long as the location of field sites is carefully planned, case studies offer important advantages over extensive approaches. This is because case studies provide an opportunity to ask fundamentally different questions in a fundamentally different way. Case studies often aim to explain the mechanisms that generate patterns observed in extensive studies (Blaut, 1959), and as such, case studies should not be judged by their representativeness (or lack thereof) but by the quality of the theoretical reasoning that they generate (Richards, 1996). Similar arguments, based on alternative views of generalization and theory validation, are used by qualitative researchers to explain the use of case studies (e.g. Miles and Huberman, 1994).

While statistical theory guides sample collection in extensive approaches, the selection of case studies for intensive study is based on less well-established criteria. Richards (1996) highlights the importance of carefully selecting cases that have those properties which facilitate rigorous tests of the hypotheses under consideration, and meticulous definition of local conditions. Curtis *et al.* (2000) review case-selection criteria formulated for qualitative research which include ethical considerations and relevance to a theoretical framework. Using examples from medical geography, they find that it can be difficult to reconcile these two criteria such that ethical considerations are at odds with the selection of the most useful or theoretically relevant cases.

The collection of information within a case study, especially in quantitative research, will typically require extensive sampling *internally*, albeit on relatively small temporal or spatial scales. If useful internal inferences are to be made then this sampling must be rigorous and statistically accountable. In turn, as Richards (1996) points out, the practical distinction between 'case studies' and 'extensive studies' is somewhat fuzzy in geomorphology, and largely semantic. In practice, many geographers are typically engaged in studies that shift between these two styles, using their complementary opportunities to explain geographical phenomena. The literature reviewed above provides a starting point for further consideration of case studies as a form of sampling. The remainder of this chapter focuses on extensive sampling and classical statistical inference.

WHAT MAKES A GOOD SAMPLE?

When sampling is used to make generalizations about a larger population, the aim of sampling is to obtain a 'representative characterization' of whichever aspects of the population one is interested in. Characteristics of a population are referred to as parameters, while those of a sample are called statistics. A good sample may then be defined as one that satisfies two criteria: it must provide an unbiased estimate of the parameter of interest, and it must provide a precise estimate of the parameter of interest.

Accuracy, precision and bias

The very nature of sampling means that, in repeated sampling of the same population, different items (individuals, businesses, households, etc.) will be drawn. Thus, statistics vary between samples and, in each case, differ to a greater or lesser degree from the population parameters that they estimate. The difference between sample estimates and the true population value is referred to as the accuracy of the sample. Accuracy is gauged in terms of its two components: bias and precision.

These terms can be illustrated by representing repeated sampling of a population by a darts competition in which each dart represents a single sample estimate and the bull's eye represents the true population value (Figure 17.1). Bias refers to

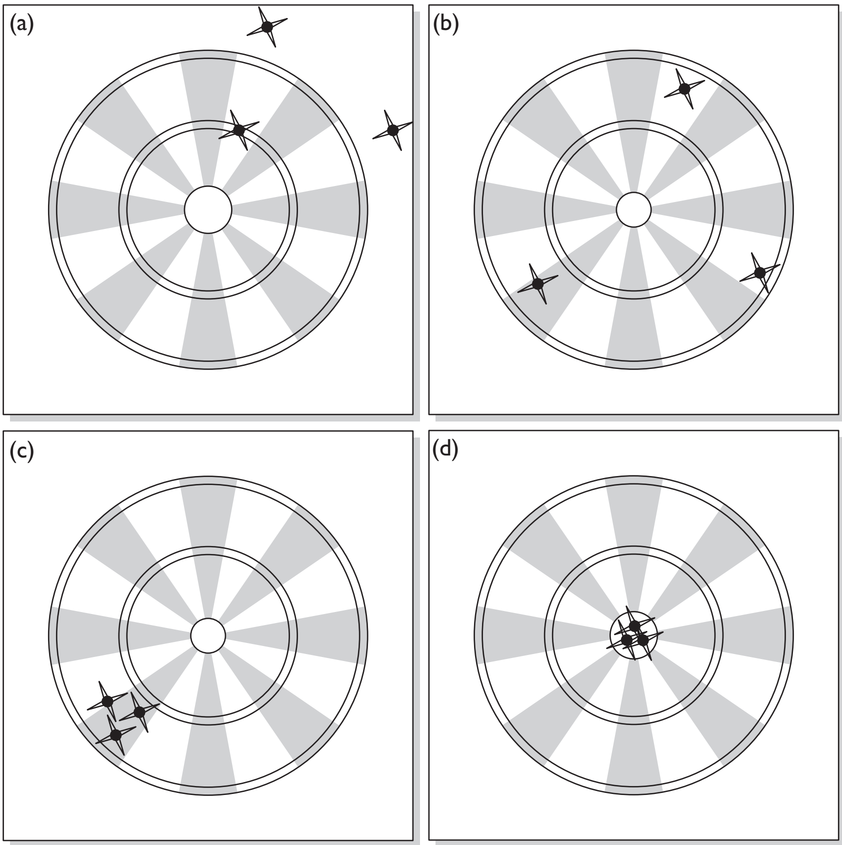

Figure 17.1 Precision and bias represented as a game of darts

the systematic deviation of sample statistics from the true value. A set of sample statistics that vary around the true value without any discernible pattern is said to be unbiased (Figure 17.1b and d) whereas a tendency to be consistently different (for example too high or too low) reveals bias (Figure 17.1a and c). Lack of bias ensures the representativeness of a sample and is a fundamental requirement of statistical inference. In principle, lack of bias is achieved by sampling randomly from a population and in practice this becomes the main challenge in sample collection. Precision refers to the size of the deviations between repeated estimates of a given statistic. It describes the degree to which repeated estimates are clustered together: are they tightly bunched (Figure 17.1c and d) or widely dispersed (Figure 17.1a and b)? In real sampling problems, bias and precision cannot be observed directly because the value of the population parameters are unknown (otherwise there would be no need to sample) and, typically, only a single sample is collected. When sampling we should then aim to maximize precision and minimize bias so that we can have faith that the single sample we collect is accurate (Figure 17.1d).

Minimizing bias

Bias is introduced into a sample in one of three main ways. First, the actions of the person collecting the data may introduce operator bias. Two different individuals asked to sample the same population may produce samples that are not just different (this is expected), but *systematically* different from one another. For example, in approaching passers-by to interview on a city street, an operator may knowingly or inadvertently select individuals of a certain age, gender or ethnicity, thereby over-representing those groups in the sample. In collecting pebbles from a beach to determine their average size, an operator may tend to pick up pebbles that are more distinctive in colour and more easily seen. Since colour depends on lithology and lithology affects size, the sample could be unrepresentative of the range of sizes present. Operator bias is reduced by careful training and adherence to a standardized set of procedural rules consistent with the sampling method (see below). Thus, the street interviewer may be asked to minimize selection-based bias by approaching every tenth person who passes by and the pebble picker may be instructed to collect the pebble that lies beneath the metre marks on a measuring tape laid down on the beach surface. However, the propensity for humans to ignore rules or simply to make mistakes means that operator bias is difficult to remove entirely. Thus, looking down at the correct spot on the tape, the pebble picker can choose between that pebble that lies to the right or left of the tape and it is common for operators to preferentially select pebbles that fit comfortably in their hand, potentially ignoring large or small clasts and thereby biasing the sample (see Marcus *et al.*, 1995 for a fuller discussion of this case). Additional procedural rules can be made to try to minimize such errors, but ultimately it may be difficult to eradicate them entirely. The potential for operator bias increases with the number of operators involved so that minimizing the number of operators helps to reduce this source of inaccuracy.

Second, bias can be introduced by faulty or misused measurement devices that systematically misrepresent the characteristic of interest – note that this is different from non-systematic measurement error discussed below. This is fairly clear in the case of instruments that measure some chemical, physical or biological property and have the potential to be mis-calibrated or broken. Mis-calibration is a problem because many measurement devices do not measure the property of interest directly but via some more easily assessed quality that is quantitatively related to the target property. For example, stream velocity is routinely measured using devices that generate a small magnetic field in the flowing water and detect changes in the electrical current that is produced as water flows through it. The current is proportional to the water's velocity, so that velocity can be determined by measuring the output current. However, care must be taken to correctly calibrate and set-up such instruments to ensure that no systematic error produces consistently deviant velocity estimates. In the widest sense, a question on a questionnaire, for example, is also a type of measurement device that can return a biased response by asking leading or biased questions (see Chapter 6).

Third, and perhaps most commonly, bias can be introduced during the design of the sampling programme, particularly poor definition of the population, or choice of an inappropriate sampling method (the method by which individual sample items are drawn from the population). In geographical research the content of the population is apt to vary in space, in time and with the scale of interest, often in a systematic way (see Chapter 24). Unfortunately, this means that it is relatively easy to collect a sample that is not representative of the intended population (these issues are discussed at length below).

Maximizing precision and minimizing non-systematic measurement errors

Precision is largely a function of three things: the number of observations that make up a sample; the heterogeneity (variability) of the characteristic of interest within the population; and non-systematic errors that arise from the technical limitations of the measuring procedure. We shall return to the first two below, suffice to say here that the larger the sample size and lower the population heterogeneity, the more precise sample estimates will be.

Measurement errors, sometimes called pure errors, exist in all measurement operations because of technical limitations within the measuring system (the people and instruments involved). These are the irreducible errors that one has to be willing to accept in making any set of observations. Wherever possible, every effort should be made to minimize them and, of course, one must endeavour to ensure that the errors are free of bias. A brief example from geomorphology illustrates several relevant issues.

In the field, we can use an instrument called a total station to survey topography – for example, exposed gravel bars in an alpine river channel. The instrument is set up at a base station and measures distances, declinations from horizontal and directional angles to survey points across the bar surfaces. A target prism attached to a pole of known length is used to mark each survey position and its reflective properties allow distance and angle measurements to be made to it. These measurements are used with simple trigonometry to derive the elevation of each point and its position within a Cartesian coordinate system. The instrument measures declinations from horizontal with an accuracy of ±5 seconds (where 1 second is 1/3600th of a degree), and distances of over 1 km with an accuracy of ±3 mm. This kind of instrument is expensive and the measurements it makes are incredibly refined. Nevertheless a number of measurement errors can be identified:

- As indicated by the manufacturer's specifications, repeated measurements of exactly the same target position return declination values that vary by as much as 5 seconds either side of the true value. In practice this error is very small, introducing a deviation of no more than 5 mm into the calculation of the elevation of a position 1 km away.
- Holding the target pole still, especially in cold, windy, weather is difficult. Despite every effort it is common for the target to move a few millimetres back and forth over a period of seconds. In turn, repeated measurements with the pole in the same position will yield very slightly different distance estimates.

- The survey aims to characterize bar topography but our measurements are affected by the smaller scale, gravelly surface texture of the topography being surveyed. Thus, choosing to place the target pole 1 cm to the left or 1 cm to the right of a particular spot can mean measuring the elevation of a hole between two pebbles or the elevation of the top of a pebble, in which case our estimate of bar elevation at that position can vary by tens of millimetres.

In practice, the first two sources of error are of little concern because they introduce only a very small amount of uncertainty into the results. They could be reduced further, for example, by moving the base station closer to the survey positions and using a tripod rather than a cold assistant to hold the target pole. The third error is more worrying because it is slightly larger. However, there is no bias involved because the assistant selecting each target position is instructed to place the pole randomly rather than, for example, consistently selecting pebble tops. This ensures that while individual points may be a few centimetres higher or lower than the average bed elevation in their vicinity, the overall surveyed surface is neither consistently above, nor consistently below the average. Most importantly, all of the errors combined are very small relative to the variations in bar topography that we aim to characterize (millimetres compared with metres). This means that we can be confident that what little uncertainty the errors do introduce does not affect our ability to describe the overall shape of the bar surfaces. Clearly, this kind of decision-making is dependent on the job at hand, its aims and whether or not the overall precision achieved is sufficient to meet the objectives. Making such a decision always depends on appreciating the measurement errors involved and, therefore, that every effort should be made to characterize them.

An alternative example from ecology illustrates how small amounts of imprecision can be important and can lead to scientific misunderstanding. It was recently reported that, in addition to growing longer as they grow older, Galapagos marine iguanas shrink during periods of reduced food availability, for example in response to major wet episodes like El Niño climatic events (Wikelski and Thom, 2000). These arguments were based on catch and release studies in which many individually identifiable animals were periodically recaptured and measured over a number of years. Shrinkage involved more than simply losing weight, with animals becoming shorter by up to 20 per cent, possibly by bone absorption. This ‘bidirectional’ growth phenomena is highly unusual, so several researchers set out to see whether it was present in other types of reptile, including snakes. Examination of a large catch and release dataset for 16 species of West African snakes revealed shrinkage in up to 6 per cent of cases for some species (Luiselli, 2005). However, careful examination of these data and the measurement errors within them revealed that the proportion of shrinkage cases was strongly correlated with the ‘measurability’ of different species. Large, vigorous, aggressive and highly venomous snakes are understandably more difficult to measure than smaller, docile, non-biting, non-poisonous snakes and there was a clear correlation in the dataset between greater handling difficulty and a higher incidence of apparent shrinkage. The implication of this is that shrinkage is not real, but an artefact of measurement error, and Luiselli (2005) concludes that snakes do not shrink but that snakes which are difficult to handle have a higher probability of being measured incorrectly so that there is a greater probability that they will appear to shrink (either because their length was overestimated to begin with or underestimated during a subsequent measurement). Although the individual measurement errors are unbiased (measurements are equally likely to be too long or too short) they are imprecise and irreducible, in the sense that it would take substantially more effort than it is feasible to expend to obtain more accurate measurements. This lack of precision is important in this case because it could have led to unreasonable conclusions (snakes shrink) if it had not been for the additional, careful analysis. This example highlights the importance of always making every effort to understand and characterize measurement errors during a sampling campaign.

DESIGNING A SAMPLING PROGRAMME

In any project there are two main controls on the design of the sampling programme: the research aims and the resources available (time, money, person power). While the research objectives should drive the sampling design, more often than not it is resource issues that limit the sampling programme, and compromises have to be made. Two key issues are the definition of the population of interest and the choice of sampling method.

Defining the target population

Defining the target population is a critical step, and begins with a clear definition of the unit of study (the items about which generalizations are to be made and that will be sampled). In social geography this might be an individual, a household, or an organization. In fluvial geomorphology it might be a channel cross-section, a bar, a hydrological link or a river basin. One of the things that makes geography such a fascinating discipline and makes sampling necessary is that the character of these units and therefore the content of the population is apt to vary, often systematically, in space, time and with the scale of interest. The population must be defined with this heterogeneity in mind, while at once satisfying the needs of the research aims and working within resource limitations. Defining the population is then an iterative process in which several questions are asked: How do the population characteristics of interest vary spatially and temporally? Which variations are important for the study and which are not? How can important sources of variation be included and unimportant sources of variation excluded? Can the research aims be modified to accommodate practical difficulties? Answering these questions depends on careful investigation of published research, consideration of what one might reasonably expect and a clear understanding of the research aims. In turn, the spatial and temporal character of the intended population should be stated and used to guide the design of the sampling programme.

Failing to accommodate temporal and spatial variability, by targeting too narrow a slice of the possible population, will produce a sample that is unrepresentative. For example, questioning households in only one enumeration district about their leisure activities is unlikely to yield results that are applicable to the city as a whole because one district is unlikely to encapsulate the range of economic, ethnic and age-related factors that influence use of leisure time across the city. Similarly, sampling suspended sediment concentration in a stream only during the rising limb of a flood is likely to yield an average value that is too high for the flood as a whole because of temporal variations in sediment availability over the course of the event. Equally, it is possible to target too much of the possible population in terms of its spatial extent, temporal boundaries or internal structures. This not only spreads precious resources thinly with implications for sample precision (see below), but may also add sources of variability that are not of direct interest and that obfuscate or dilute critical information. Thus, it is often necessary to exclude sources of variability from a sampling programme and focus attention on particular objects, places, times or patterns.

Choice of sampling method

Having clarified the spatial, temporal and structural dimensions of the target population, the next problem is to determine the best way of sampling from this target population. A variety of sampling methods are used by geographers and fall into two basic groups: non-probability methods and probability-based methods. Non-probability methods cannot be used to make statistical inferences about the population from which they are drawn. In choosing to adopt non-probability methods one must therefore accept that statistically rigorous representativeness is not a primary issue in the research design (which may be the case, for example, in some research utilizing case studies). If the intention is to make generalizations about a larger population then non-probability methods should only be used with extreme caution and it is in this context that such methods are briefly reviewed here.

In *accessibility sampling*, units are selected on the basis of convenience, such that one selects the most accessible units from the population. Such samples are likely to yield a biased sample. An example from biogeography illustrates the potentially serious consequences if this sampling flaw remains unrecognized. Reddy and Dávalos (2003) examined the spatial distribution of 3,504 sites in sub-Saharan Africa where passerine (perching) bird species were observed between the 1800s and 1970. Datasets of animal distributions like this, compiled in a large number of studies over many years, are important because they provide information over large areas that are used to define biodiversity hotspots and priority locations for conservation. What Reddy and Dávalos (2003) found, however, was that the location of sampling was strongly influenced by accessibility, with sampling sites concentrated in a non-random pattern close to cities, roads and rivers (Figure 17.2). The use of these data to identify conservation priorities is, therefore, problematic because the information is biased. It may be that some of the apparent hotspots targeted for conservation are less rich in species than other locations that are less accessible (a long way from roads, rivers and cities) where little or no information exists. The implications for conservation biogeography are that greater effort is required to collect information in less accessible locations and to develop methods for correcting accessibility bias where such fieldwork is impracticable.

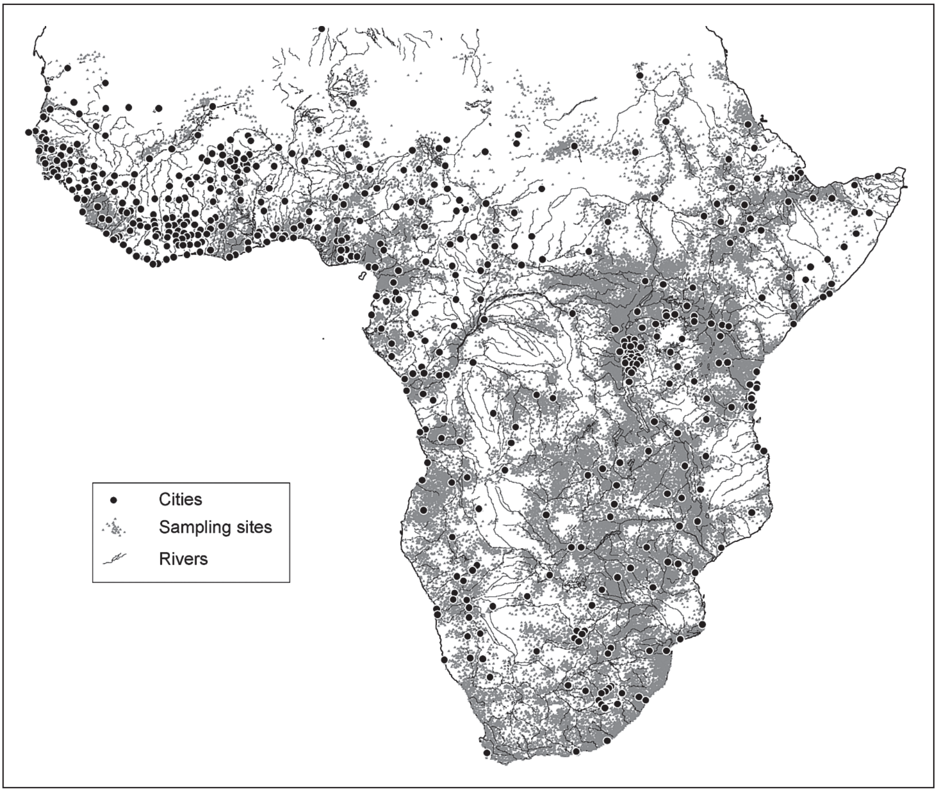

Figure 17.2 Map of sub-Saharan Africa showing approximately 3,504 locations where sampling has been conducted for passerine birds between the 1800s and 1970 (light grey dots). Major rivers (dark grey lines) and cities (large black dots) are also shown. In many regions, sampling locations tend to be located relatively close to cities and rivers; a pattern that is confirmed using formal testing. This illustrates accessibility bias in the selection of sampling sites. That is, sites close to rivers and cities are over-represented because they are relatively easy to access. See Reddy and Davalos (2003) for a full discussion and details of the data sources

In *judgemental* (also referred to as *purposive*) *sampling*, units are selected subjectively by the researcher on the basis of prior experience. This is problematic because the researcher's previous experience may be limited and his or her own prejudices, derived from his or her expectations and viewpoint, become an integral part of the selection process. Even if, by chance or skill, a judgemental approach yields an unbiased sample, it is difficult to prove that this is the case and therefore difficult to convince critics of the value of any generalizations that are made. *Quota sampling* aims to be more representative by attempting to produce a sample that replicates the general structure of the population. Predefined quotas based on factors like age, gender and class are filled, thereby imposing some useful control on the selection of units, but the choice of individual items within each quota group is still subjective. Kitchin and Tate (2000) suggest that this method can yield representative samples but should only be used in situations where prior work has shown this to be the case.

In contrast, *probability-based sampling* methods aim to preclude bias and produce representative samples. Their common characteristic is that the sampling units are selected by chance and the probability of any unit being selected can be determined. Probability-based methods must be used if one intends to use inferential statistics to generalize from the sample to the population. These methods require that a sampling frame exists or can be developed. A sampling frame is a list or other representation of the target population from which units can be drawn (for example an electoral roll, a catalogue of discharge gauging stations, an aerial photograph, a map, or a street directory).

Table 17.1 illustrates several probability-based methods. The two basic methods are *simple random* and *systematic sampling*. Their common feature is that there is an equal probability of selecting each and every unit within the sampling frame. Two issues are worth considering when adopting these methods. First, if systematic sampling is applied within a sampling frame that includes a repetitive structure and the sampling interval that is chosen coincides with that structure, then bias will be introduced. For example, many alluvial rivers exhibit repetitive pool-riffle-bar morphology in which the spacing between units is typically five to seven times the channel width. If water depth or grain size or stream velocity are systematically sampled using a similar interval, it is possible that measurements will be biased toward the characteristics of pools or riffles. Second, with a target population where the characteristic of interest is heterogeneous but also exhibits some internal pattern, it is important to obtain uniform coverage of the sampling frame without any gaps. Simple random sampling may not do this as well as systematic sampling because it is possible for sampled units to be unevenly distributed, as illustrated for the case of river sediment characterization by Wolcott and Church (1991).

In a *stratified sample* a number of homogeneous sub-groups or strata, differentiated by some relevant characteristic, are recognized within the population. In contrast to the simple and systematic methods, the probability of selecting an individual unit from the sampling frame varies, depending upon the stratum that the unit belongs to. Three common reasons for utilizing stratified sampling illustrate its value. First, it can be used to ensure that the number of units drawn from distinctive strata is in proportion to their true size in the population. This is known as *proportionate stratified sampling*. Simple random and systematic sampling will achieve this by default if the sampling frame is appropriate, comprehensive and accurate, which should be the case if the sampling frame is developed for the research project. However, it is not uncommon for the sampling frame to be obtained from a source that compiled the frame for purposes other than those for which it is now intended. Such frames may be biased in favour of one or other strata. Similarly, instrument malfunction at a particular time or place, or non-responses to questionnaire surveys may yield a sample that is known to be biased. In either case, if the true population proportions are known then each strata can be randomly sub-sampled in those proportions to obtain an unbiased sample. Second, it may be uneconomical or unfeasible to sample strata of very different

Table 17.1 Basic sampling methods

	Description	Physical illustration	Human illustration
		Sediment size *Aim*: ascertain average size of the sediment particles on a river bar. *Population*: all particles on the river bar. *Unit*: a sediment particle. *Frame*: a map of the bar surface located in an arbitrary cartesian space. *Measurement*: using a size template.	**Street safety** *Aim*: ascertain views of university students on campus safety. *Population*: all students at the university. *Unit*: an individual student. *Frame*: a list of students and their addresses. *Measurement*: by questionnaire.
a) Simple Random	Within the sampling frame each unit is assigned a unique number or position. Numbers and thence units are selected at random from the sampling frame.	A random number generator is used to pick x and y coordinates. These coordinates locate particles for measurement.	Each student on the list is assigned a unique number. A random number generator is used to pick numbers and the corresponding people are sent questionnaires.
b) Systematic	A sampling interval is defined (e.g. every 10 m, every fourth individual, every 60th second). The first unit is randomly selected as in (a) and subsequent units are selected systematically according to the sampling interval.	The bar is approximately $40\,m^2$ and a sample of 100 is required. A sampling interval of 2 m is defined. From an arbitrary origin, a grid of 2 m squares is projected onto the sampling frame map. Grid intersections locate particles for measurement.	The list contains 500 names and a sample of 100 students is required. An interval of 4 units is defined. One name is randomly selected as above. Subsequently, every fourth student is selected. If the end of the list is reached, counting continues at the beginning.
c) Stratified	Mutually exclusive sub-groups (strata) are identified and sampled randomly or systematically in one of two ways: *Proportionate*: each stratum is sampled in proportion to its true population proportion. This is necessary if the sampling frame is inadequate. *Disproportionate*: an equal number of units are sampled from each stratum irrespective of their true population proportion. This is necessary when comparisons between strata are required.	Four strata corresponding to distinct facies (areas of homogeneous sedimentary character) are evident. In this case the frame is adequate and there have been no measurement problems. Simple random and systematic sampling are adequate. Suppose one wishes to compare size in facies 1 (a small area) and facies 4 (a larger area). An equal number of particles should be selected from each. Thus, disproportionate sampling is necessary (note this will yield a biased sample of the population so weighting is required).	It is suspected that gender is an important factor in determining views on campus safety. Suppose the supplied list is for students in only one faculty. Different faculties typically exhibit distinct gender distributions. In this case the list is not representative of gender distribution across the university. Proportionate sampling is required: stratify (male, female) and randomly sample in each group to obtain numbers that yield the female:male ratio for the university as a whole. .

sizes in proportion to their size (total area, number of units, etc.). A more efficient method is often to collect a random sample of common size within each stratum, then weight the statistics obtained for each strata according to the stratum's size within the population, and combine them appropriately in order to generate population estimates. Sampling the same number of units from strata of different size is referred to as *disproportionate stratified sampling*. Third, individual research projects may ask questions about the strata, often requiring that comparisons are made between them. In this case it is necessary to obtain equally precise samples for each stratum, which means selecting a similar number of units from each. Simple random or systematic sampling does not do this, but rather selects a number of units from each stratum that is in proportion to the stratum's size. With disproportionate stratified sampling this problem is overcome by randomly selecting the same number of units from each stratum, irrespective of their true relative sizes. In using this method it is important to remember that as far as the population as a whole is concerned, one has created a biased sample so that if estimates of population parameters are required, strata estimates must be combined using appropriate weighting techniques.

A final example of a probability-based method is the *multi-stage* or *hierarchical sample* in which the sample is selected in several stages that usually relate to spatial or temporal scale. For example, if the campus safety study (Table 17.1) was extended to a global scale the aim might be to sample 100 universities from around the world. First, ten countries might be randomly selected, then within each country five cities, and ultimately within each city, two universities. Multi-stage surveys are an efficient method when faced with a very large population in space or time.

Choosing between these various probability-based methods (and the many others that have been suggested) requires some prior knowledge or reasoned judgement concerning any spatial or temporal structures within the population, a thorough understanding of the sampling frame and a clear set of aims. Without a good appreciation of these it is possible to inadvertently choose a sampling method that systematically favours some parts of the population over others, in which case the characteristics of interest are not properly represented. This basic point has been stressed by several authors who have considered the specific details of applying standard sampling methods to spatial data (e.g. Berry and Baker, 1968; Harvey, 1969). Haining (1990) suggests that systematic sampling is superior where the underlying spatial variation is random. Wolcott and Church (1991) find that a particular combination of grid and random sampling known as stratified systematic unaligned sampling (cf. Smartt and Grainger, 1974; Taylor, 1977) performs well for areally structured data. They point out that it avoids the primary problems with each of random and systematic sampling: the possibility that random sampling is unevenly distributed thereby missing small spatial structures, or that the data contain spatial structures that have the same spacing as the grid spacing, thereby introducing bias.

In summary, non-probability methods are less desirable than their probability-based counterparts and certain probability methods are more appropriate than others in certain circumstances. Nevertheless, it is important to recognize that the vagaries of empirical research often make meeting the ideal difficult (if not impossible) with the result that the target population and the sampled population differ

(Krumbein and Graybill, 1965). This might be because the resources necessary are not forthcoming. It may be that accurate information about the population is not available to guide programme design or that there are unknown and hidden sources of variation within the population. It may be that an appropriate sampling frame does not exist or that we are forced to accept an accessibility sample because only certain people will talk with us or only certain places can be reached. In cases like these it is incumbent on the researcher to make it very clear exactly how sampling was conducted and for him or her to interpret his or her results in light of suspected sampling weaknesses.

Analytical requirements

Finally, in designing a sampling programme it is important to think ahead to the analytical stage of the research and identify any restrictions or requirements that the intended analysis imposes on the sampling strategy. For example, it may be that the inferential statistics used require a minimum number of samples or that a laboratory machine requires individual samples to be of a particular mass. It is certainly the case that any hypothesis being tested will require the data to be collected in a particular manner. In experimental and some observational projects the experimental design will be an integral part of designing the correct sampling programme. It is therefore crucial to identify the analytical procedures that will be used in the laboratory or at a desk *before* setting out with clipboard or shovel.

STATISTICAL INFERENCE

We have already noted that geographical enquiry must deal with uncertainties. Hicks (1982: 15) defines inferential statistics as ‘a tool for decision making in the light of uncertainty’, and geographers have certainly found inferential statistics to be a valuable tool. Inferential statistics use sample data to make probabilistic statements about the population from which they are drawn. Statements can be made about the characteristics of the population, which is referred to as parameter estimation, and also whether a particular supposition about the population is true or false, which is referred to as hypothesis or significance testing.

Numerous text books are available that explain the principles and practical application of the great array of inferential statistical techniques used by geographers. These include specifically spatial techniques that extend statistical analysis to the examination of patterns in space (e.g. Norcliffe, 1977; Williams, 1984; Haining, 1990; Shaw and Wheeler, 1994; Fotheringham *et al.*, 2000; Rogerson, 2006). Particular attention should always be paid to the data assumptions that these procedures have and whether so-called parametric or non-parametric techniques are most appropriate. There are also some specifically geographical issues to be aware of too, particularly spatial autocorrelation. This refers to the propensity for the value of a variable at one location to be related to the value of that same variable in a nearby location. It is problematic, because inferential statistical techniques often require that each sample measurement is independent of all others. In spatial data, autocorrelation is common (otherwise location would not matter) such that the performance of standard methods may be degraded and there is the potential for misinterpretation. It is possible to measure the significance of spatial autocorrelation in a data set (see, for example, Kitchen and Tate, 2000) and standard inferential procedures can be adapted to minimize its impact (see, for example, Cliff and Ord, 1975; Fotheringham *et al.*, 2000; Rogerson, 2006). Chapter 26 provides alternative means of describing and exploring spatial associations.

There are many introductory texts that can provide a detailed step-by-step introduction to inferential statistical methods. The aims of this section are limited to explaining the apparent incongruity of statistical inference – how can one make statements about a population based on a single sample drawn from it, even though one knows that no two samples would ever be exactly the same? – an apparent leap of faith that brings to mind Jean Baudrillard's comment that, 'Like dreams, statistics are a form of wish fulfilment' (Baudrillard, 1990: 147). The simple answer is that, although we know our sample to be unique, statistical theory allows us to assess the reliability of sample estimates (called statistics) such as the sample mean. It is, therefore, possible to ascertain the likely difference between a sample statistic and the equivalent population parameter *without knowing* the value of the population parameter. In turn, the differences between sample statistics, for example mean values from different groups, can be compared with one another to test the hypothesis that they come from different populations. The following exposition of these ideas is necessarily very brief and non-technical and focuses on ascertaining reliability rather than hypothesis testing. The reader is directed to one of the above texts (Shaw and Wheeler, 1994; Rogerson, 2006) for a fuller account.

Probability and the 'normal' distribution

A basic understanding of probability distributions is necessary before continuing. A probability distribution describes the changing frequency with which particular values of a variable of interest are measured. It is commonly visualized as a histogram in which the ordinate shows the number of occurrences (the frequency) with which groups of values occur. For example, one can describe the frequency distribution of beach pebble sizes in a 100-pebble sample by indicating the number of particles in each of several consecutive 10 mm grain-size classes. Frequencies can be represented as absolute numbers or as relative proportions, in which case they represent the empirical probabilities of measuring a value in each class. Thus, if 35 of the 100 pebbles were found to be between 40 and 50mm in diameter, it follows that there was a probability of 0.35 (a 35 per cent chance) of finding a pebble in that size range on the beach. Probability distributions for measured phenomena take a wide variety of forms, but a typical situation is that values close to the mean are common and those further away are proportionately less common. Specifically, many phenomena exhibit an approximately 'normal' distribution (sometimes referred to as a Gaussian distribution after the mathematician who first defined it) with its characteristic bell-shaped curve, centred on the mean.

The properties of the normal distribution, and in turn empirical probability distributions that approximate it, are at the heart of basic statistical inference. Any normal distribution can be described in a standardized form in which raw empirical data are transformed into so-called z-values. These numbers express changes in the measured values as multiples of the data's standard deviation. In standardized form the mean of the distribution is zero and the standard deviation of the distribution is 1.0. Because the mathematical form of the standardized distribution is known, the probabilities of observations falling within any given range of z values can be calculated and most statistical text books contain tables that give the probability associated with specific z ranges. Thus, there is a 0.68 probability (0.34 either side of the mean) of a standardized observed value falling in the range z = −1.0 to z = 1.0; i.e. within one standard deviation of the mean. Similarly, 95.45 per cent of observations will be within two standard deviations and 99.73 per cent within three standard deviations of the mean. This is true of any normally distributed variable which means that we can apply such reasoning to a wide variety of phenomena in physical and human geography. By using such tables in reverse, the values of z that are associated with selected probabilities can be ascertained. For example, 95 per cent of the values (47.5 per cent either side of the mean) in a normally distributed phenomenon will have z values that are in the range ±1.96. Equally, sampled values with z values outside this range have a probability of being observed 5 per cent of the time or less.

Confidence statements about sample statistics

If repeated samples are drawn from the same population and in each case the mean is calculated, the mean values will vary from sample to sample but will tend to cluster around the true mean of the population. Such a collection of sample means (or indeed any other sample statistic) is called a sampling distribution. A piece of mathematical theory called the Central Limit Theorem (CLT) proves that sampling distributions are normal with a mean value equal to the value of the true population parameter (e.g. the true population mean) and that this holds irrespective of the population distribution. Thus, even for a phenomenon that does not exhibit a normal distribution, the sampling distribution of the mean is normal. The standard deviation of a sampling distribution is known as the standard error and it has the same general properties as the standard deviation of any normal distribution so that, for example, 95 per cent of the sampling distribution lies within 1.96 standard errors of the true population mean.

Standard errors can be determined empirically by repeated sampling of a given population, but this is rarely plausible. It is of significant consequence, then, that standard errors can be calculated on the basis of collecting only a single sample. For example, the standard error of sample means (σ_x) can be calculated as

$$\sigma_x = s / \sqrt{n}$$

where s is the standard deviation determined from a single sample and n is sample size. Armed with this value and our knowledge of the normal distribution it is possible to make statements about the reliability of the sample mean; that is, to say how confident we are that the true population mean is within a given interval about the sample mean. Remembering that the probabilities in a z table indicate that 95 per cent of a normal distribution lies within 1.96 standard deviations of the mean, we can say that there is a 95 per cent chance that the sample mean lies within 1.96 standard errors of the true population mean. This is equivalent to saying that there is a 95 per cent chance that the population mean lies within 1.96 standard errors of the sample mean.

For a given case, the interval can be specified in the original data units and is known as a confidence interval. So, for example, for a sample of pebble diameters with a standard deviation of 20 mm and n = 100, there is a 95 per cent chance that the sample mean lies within $1.96 \times \sigma_x = 1.96 \text{ x } (20/\sqrt{100}) = \pm 3.9$ mm of the population mean. This is commonly interpreted as meaning that in 95 samples out of 100 the sample mean would lie within ±3.9 mm of the population mean, although more precisely it says that if 100 samples were used to construct 100 confidence intervals the true population mean would be included within 95 of them. Confidence intervals for any probability can be constructed using the appropriate z value, so that at 0.99 probability the confidence intervals in the above example are $2.58 \times (20/\sqrt{100}) = \pm 5.2$ mm. An important caveat for the reader to investigate further is that while large samples always have normal sampling distributions, irrespective of the population distribution, small samples ($n < 30$) have distorted distributions with a form that is a little different from 'normal'. Small samples tend to yield statistics that are distributed according to the t-distribution, sometimes known as 'Student's t-distribution'. This has similar properties to the normal distribution and it is used in the same way to determine the reliability of sample estimates, except that probability values from published t-tables, rather than z-tables, are used.

The CLT and standard errors are so important because they allow us to make rigorous statements about the reliability of the statistics we derive from sample data – that is, to accurately quantify the uncertainty that is inherent in a sample. In turn, they provide a basis for making rigorous comparisons between samples and thence for testing hypotheses. Just as confidence intervals are used in assessing reliability, so-called significance levels, denoted by α, are used to attach probability statements to the decisions made in hypothesis testing. There is always the chance that a given decision is incorrect and levels of significance define the probability that one incorrectly rejects a true hypothesis. Significance levels are set by the researcher as part of the testing procedure. Usually, we are only willing to accept low-levels of error, so significance levels are set to 5 or 1 per cent, though smaller, more stringent values can be used. The important point to make at the end of this section is that statistical inference, beyond the mathematical formulation of the various procedures and tests, involves commonplace ideas of confidence and significance *not* certainty. It allows us to attach probability statements to estimates and decisions but crucially, statistical techniques do not provide binary, 'black and white', yes and no answers. It is always the responsibility of the researcher to choose levels of confidence and significance, and to interpret results thoughtfully in light of these choices.

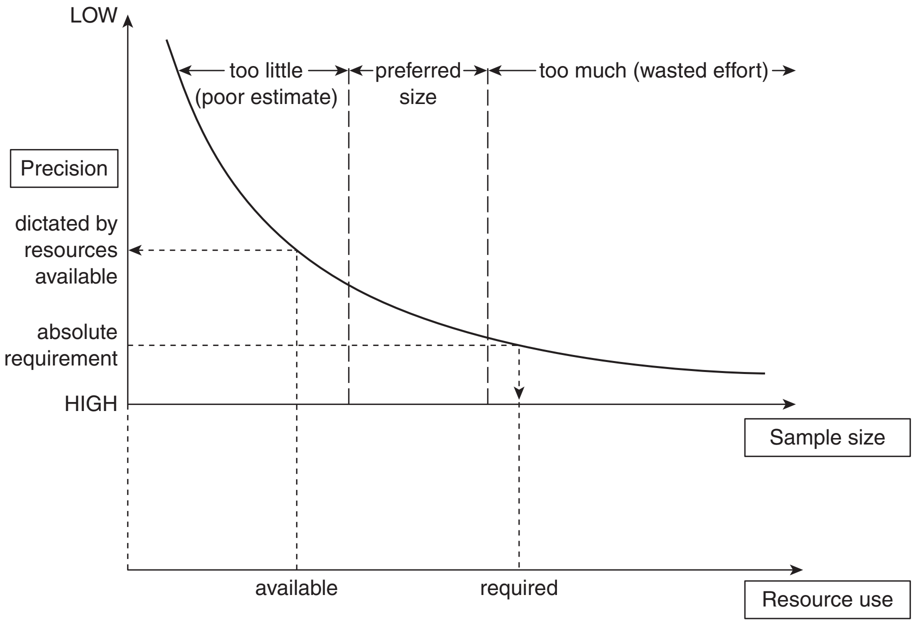

Figure 17.3 The relation between estimate precision, sample size and sampling resources

SAMPLE SIZE

A frequently asked question is, 'How big should my sample be?' The answer reflects a compromise between the desired precision of the sample estimates and the resources available, because maximizing precision and thence the significance that can be attached to statistical inferences (see above) requires the collection of large samples, but also demands greater resource expenditure.

In general, precision improves with sample size in a curvilinear fashion. As sample size increases precision improves rapidly to begin with but then more slowly (Figure 17.3). In this case, we can imagine that in any sampling procedure there is an optimal sample size corresponding to that region where the curve in Figure 17.3 begins to flatten out. Beyond this point additional gains in precision are small and do not warrant the additional sampling effort (resources) required. Before this point, the sample size is too small to yield reasonable estimates of the population characteristics and it is worth expending small amounts of effort to obtain significantly better estimates. Ideally, we want to obtain samples of a size somewhere in this optimal zone. Several ways of doing this are discussed below, but it is first worth noting two cases where the option of seeking the optimal solution does not arise.

First, a common situation is that the resources needed to collect a sample of the ideal size are simply not available so that low sample precision is inevitable. Similarly, where secondary rather than primary data is being used, the secondary data may be less voluminous than that desired. When designing a sampling programme it is always necessary to carefully consider the allocation of resources in light of this problem. It may be that sacrifices can be made in one part of the programme in order to improve precision elsewhere. For example, rather than obtaining low-precision estimates of water quality in twenty lakes, it may be advisable to seek quality estimates for five lakes, especially if the retained lakes are carefully selected to test one or more hypotheses.

Second, it may be that a specified level of precision is required by the research aims in which case there may be no option but to collect inefficient, large samples. Work with legal or health implications often has an absolute level of precision that is required by the project objectives. A related issue that constrains sample size is the possibility that a particular physical or statistical technique used to analyse the sample data has sample size requirements. It is therefore important to know how your data will be analysed before the sampling programme is finalized.

In any given study, the relation between sample size and precision is driven by the variability (heterogeneity) of the characteristic of interest within the population. For a given sample size, precision is worse in populations that exhibit greater spread or variability and the more heterogeneous a population, the greater the sample size required to obtain a given level of precision.

Formulae exist for calculating the sample sizes needed to obtain specified levels of precision for a given statistic. For example, in the case of estimating the population mean μ, precision can be thought of as the error, δ, that we are willing to accept – that is, the acceptable difference between a sample mean, x and μ ($\pm\ \delta$ units). δ is equivalent to half of a confidence interval and a confidence interval has length $2.(z.\sigma_x)$, where σ_x is the standard error of the mean and z is the tabled value associated with the chosen significance level α. Thus,

$$\delta = z.\sigma_x$$

therefore:

$$\delta = z.(s / \sqrt{n})$$

where s is the sample standard deviation, and

$$n = (z^2.s^2) / \delta^2$$

This gives the sample size n, needed to obtain an estimate of the mean that is within δ units of the population mean with a $100.(1–\alpha)$ per cent level of confidence.

Similar formulae can be developed for estimating other statistics or for use in hypothesis testing. A device known as an Operating Characteristic Curve can also be used to determine optimal sample sizes in hypothesis testing. The operational problem with these methods is quite simply that usually we do not know the sample standard deviation beforehand. This can be overcome by a two-phase sampling procedure or by estimating the standard deviation from previously published work. An additional problem is that researchers seldom find it easy to define an acceptable error, δ.

Empirical approaches to sample size determination may then be useful. As sample size increases from one, the value of any statistic will vary significantly as successive population values are added, but will gradually achieve a degree of stability. This indicates that the sample has incorporated most of the variance evident within the population (see again Figure 17.2). If it is possible to monitor the value of the statistic of interest as the sample is collected, sampling can be curtailed when values for successive n become relatively stable. This can be an especially effective method if the same type of sample is to be obtained from a number of strata or discrete sampling frames where it is anticipated that there is little change in the population variance between those strata or frames. A pilot exercise conducted in one case can then be used to inform sample size for the whole programme. For example, in the case of an insect survey consisting of many discrete quadrat samples where the aim is to examine variations in number of taxa present, it may be worthwhile to conduct a pilot exercise in which one monitors the changing number of taxa as the size of the quadrat is gradually increased. A graph can be plotted of area against number of taxa and the stabilization point will reveal the optimal quadrat size, to be utilized throughout the survey, for obtaining a reasonable estimate of taxa number (e.g. Chutter and Noble, 1966; Elliot, 1977: 128). More elaborate empirical methods can also be used to examine sample precision and identify optimal sample sizes, for example a technique called bootstrapping (e.g. Rice and Church, 1996), though these require very large data sets and the ability to invest resources in a significant pilot study.

CONCLUSION

A principal aim of most geographical research is to make useful generalizations that might help to model or otherwise understand better the uncertain human and physical words that are the geographer's realm. Because it is usually impossible or impractical to observe all instances of variation, a smaller number of instances (the sample) are used, from which the 'population' characteristics can be estimated. Achieving this in a reliable and reproducible fashion is the basis of sampling theory and sampling design.

Geographers recognize the value of both extensive statistical sampling and intensive 'case-study' sampling. A benefit of extensive sampling is that a set of techniques known as inferential statistics can be applied to make probabilistic statements about the population from which the sample is drawn. Sampling is therefore a powerful tool, but geographical research frequently engages with very heterogeneous phenomena that require careful sampling in order to maximize the accuracy of inferential conclusions. Careful design of the sampling programme is crucial and is driven by both the research aims and available resources. The overall message is that sampling is a tool that most geographers need, but if research methodology aims to be as impartial and free of error as possible, sampling must be done thoughtfully and rigorously.

Summary

The most important aspects of sampling are as follows:

- Ensuring the quality of the sample by maximizing the precision and minimizing the bias in any measurements or observations.
- Relating individual observations and sample sets to the observed or expected geographical patterns forming the population by the correct sampling design (choosing the sampling method and the sample size).
- Assessing the significance of sample estimates using graphical and statistical methods, including the use of inferential hypothesis testing.

Further reading

- Most basic textbooks on statistical analysis include an introductory section on sampling and two that are not overly technical are Shaw and Wheeler (1994) and Rogerson (2006). The latter is useful in terms of addressing autocorrelation issues.
- Kitchin and Tate (2000) is a general book on research methods in human geography that contains some useful examples of sampling schemes and a lot more besides, and can be recommended to physical geographers for its coverage of basic statistical techniques.
- Haines-Young and Petch (1986) contains a brief but lucid overview of measurement errors and statistical inference in Chapter 11. Harvey (1969) and Richards (1996) provide useful, and in the latter case advanced, discussions of the role of case studies in geography.
- For a detailed consideration of the benefits of alternate spatial sampling techniques look at Wolcott and Church (1991). Several hypothetical examples used here have been drawn from my own experience of sampling sediments. For anyone embarking on a project involving sediment sampling, Bunte and Abt (2001) provides a plethora of valuable information.

Note: Full details of the above can be found in the references list below.

References

Baudrillard, J. (1990) *Cool Memories* (English translation by Chris Turner). London: Verso.

Berry, B.J.L. and Baker, A.M. (1968) 'Geographic sampling' in B.J.L. Berry and D. Marble (eds) *Spatial Analysis: A Reader in Statistical Geography*. London: Prentice Hall.

Blaut, J.M. (1959) 'Microgeographic sampling', *Economic Geography*, 35: 79–88.

Bunte, K. and Abt, S.R. (2001) 'Sampling surface and subsurface particle-size distributions in wadable gravel- and cobble-bed streams for analysis in sediment transport, hydraulics and streambed monitoring', *General Technical Report RMRS-GTR-74*, USDA Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado.

Chutter, F.M. and Noble, R.G. (1965) 'The reliability of a method of sampling stream invertebrates', *Archive Hydrobiologia*, 62(1): 95–103.

Cliff, A. and Ord, J.K. (1975) 'The comparison of means when samples consist of spatial autocorrelated observations', *Environment and Planning A*, 7: 725–734.

Curtis, S., Gesler, W., Smith, G. and Washburn, S. (2000), 'Approaches to sampling and case selection in qualitative research: examples in the geography of health', *Social Science and Medicine*, 50: 1001–1014.

Elliot, J.M. (1977) 'Some methods for the statistical analysis of samples of benthic invertebrates', *Freshwater Biological Association, Scientific Publication* 25. Windermere: Freshwater Biological Association.

Fotheringham, A.S., Brunsdon, C. and Charlton, M. (2000) *Quantitative Geography: Perspectives on Spatial Data Analysis*. London: Sage Publications.

Haines-Young, R. and Petch, J. (1986) *Physical Geography: Its Nature and Methods*. London: Harper and Row.

Haining, R. (1990) *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge: Cambridge University Press.

Harvey, D. (1969) *Explanation in Geography*. London: Edward Arnold.

Hicks, C.R. (1982) *Fundamental Concepts in the Design of Experiments* (3rd edn). New York: CBS College Publishing.

Kitchin, R. and Tate, N.J. (2000) *Conducting Research Into Human Geography: Theory, Methodology and Practice*. Harlow: Prentice Hall.

Krumbein, W.C. and Graybill, F.A. (1965) *An Introduction to Statistical Models in Geology*. New York: McGraw-Hill.

Luiselli, L. (2005) 'Snakes don't shrink, but "shrinkage" is an almost inevitable outcome of measurement error by the experimenters', *Oikos*, 110: 199–202.

Marcus, W.A., Ladd, S. and Stoughton, J. (1995) 'Pebble counts and the role of user-dependent bias in documenting sediment distributions', *Water Resources Research*, 31(10): 2625–2631.

Miles, M. and Huberman, A. (1994) *Qualitative Data Analysis*. London: Sage.

Norcliffe, G.B. (1977) *Inferential Statistics for Geographers*. London: Hutchinson and Company.

Reddy, S. and Dávalos, L.M. (2003) 'Geographical sampling bias and its implications for conservation priorities in Africa', *Journal of Biogeography*, 30: 1719–1727.

Rice, S.P. and Church, M. (1996) 'Sampling surficial fluvial gravels: the precision of size distribution percentile estimates', *Journal of Sedimentary Research*, 66(3): 654–665.

Richards, K. (1996) 'Samples and cases: generalisation and explanation in geomorphology', in B.L. Rhoads and C.E. Thorn (eds) *The Scientific Nature of Geomorphology*. Chichester: John Wiley and Sons.

Rogerson, P.A. (2006) *Statistical Methods for Geography: A Student's Guide* (2nd edn). London: Sage Publications.

Schumm, S.A. (1991) *To Interpret the Earth: Ten Ways to be Wrong*. Cambridge: Cambridge University Press.

Shaw, G. and Wheeler, D. (1994) *Statistical Techniques in Geographical Analysis* (2nd edn). London: David Fulton Publishers.

Smartt, P.F.M. and Grainger, J.E.A. (1974) 'Sampling for vegetation survey: some aspects of the behaviour of unrestricted, restricted, and stratified techniques', *Journal of Biogeography*, 1: 193–206.

Taylor, P.J. (1977) *Qualitative Methods in Geography*. Boston: Houghton-Miffin.

Wolcott, J. and Church, M. (1991) 'Strategies for sampling spatially heterogeneous phenomena: the example of river gravels', *Journal of Sedimentary Petrology*, 61(4): 534–543.

Wikelski, M. and Thom, C. (2000) 'Marine iguanas shrink to survive El Niño', *Nature*, 403: 37–38.

Williams, R.B.G. (1984) *Introduction to Statistics for Geographers and Earth Scientists*. London: Macmillan.