CHAPTER 8

Sampling

Learning Objectives:

- What are populations, samples, and sampling frames?
- What are probability and nonprobability sampling designs?
- What are the implications of the particular sampling frame and sampling design used in a study?
- What are some implications of sampling from continuous fields instead of discrete entities?
- What should you consider when deciding on the appropriate sample size (number of cases) for a study?

In Chapter 1, we discussed scientists' preference for parsimonious explanation and, consequently, their preference for general over idiosyncratic truths. That means scientific geographers strive for truths that apply as widely as possible. But geographers run into a dilemma here. They want to be general in their conclusions, but when they conduct studies, they are necessarily limited to relatively small numbers of cases, places, times, variables, measures, and other research entities. This is often because researchers have only so much time, money, and other resources available to do their studies. But there is an even more fundamental reason for this dilemma. Even with great resources, researchers typically cannot even potentially access all the entities to which their theories and models are relevant. For example, if we test a theory of urban development, we want it to apply to all cities at all times and places, as much as is possible given existing ideas and evidence. Even if we limited our focus, as we likely would, to cities of particular size ranges or cities within particular cultural or economic systems, we would still be concerned with a very large number of cities, some that no longer exist and others that will only exist in the future. We cannot measure cities that do not exist when we conduct our study, but we want theories that apply to these cities, if possible.

As another example, we might measure microbes in soil taken at various depths. No matter how many measurements we take, we cannot access every volume of soil in a large field, let alone everywhere on the earth where that soil type exists. Nor would it be feasible to test for every possible microbe that could be in the soil. And what about soil from the past or soil that has yet to be created? Even with virtually unlimited time and effort, therefore, we cannot do a study that includes every case, place, time, variable, or measure in which we are interested.

The solution to this dilemma is sampling. **Sampling** is any way of selecting a subset from the entire set of entities of interest, called a **population**; the resulting subset is a **sample**. Because the sample is an incomplete subset, it is always necessarily smaller than the population from which it comes. Often it is much smaller. But size does not define whether a set of entities is a sample or population. That is determined by your research goals—by what you are interested in or want to generalize your research conclusions to. In basic science, researchers usually want to generalize their conclusions to very large, even hypothetically infinite, sets of entities. So even when they can access large sets of entities, the sets are typically samples. For example, researchers who study the reflective "signatures" of different earth-surface land covers detected via satellite remote sensing (Chapter 12) routinely have samples of millions of sensor readings. On the other hand, quite small sets of entities can constitute entire populations; for example, a teacher who wants to know how well the class did at learning the material taught in a particular term does not want to generalize to other students or terms or class material, so a total of two exam scores from 15 students can constitute a population of scores. Furthermore, when a researcher's goal changes, what was once considered a population may become a sample. The distinction between samples and populations can be subtle, but it is important. If we have the entire population of entities of interest, we can analyze it by simply describing it in various ways. If, as is so often the case, we have only a sample of the entities of interest, we not only describe that subset but also consider what the sample might tell us about our larger population of interest; in Chapter 9, we discuss the logic of doing this formally as part of "statistical inference."

As we suggested above, researchers sample many different types of entities, not just cases. In fact, virtually anything that plays a role in empirical research can be understood as a sample from a larger set of possible choices that could have been made. An urban geographer whose research team records graffiti while hiking along streets and alleys is sampling graffiti, cities, neighborhoods, years, streets, and alleys. Even the particular people collecting the data and the particular directions they turn their heads while looking constitute samples of some larger population of possibilities. Most discussions of sampling focus on sampling cases, as we will for most of the rest of our discussion. However, keep in mind that most of our discussion could apply to a variety of research entities other than just cases.

We also pointed out above that researchers sample because they are fundamentally limited in their ability to access all the cases they want to generalize to, but that in practice, they are usually also limited because they have only so much time, money, personnel, and so on. That is, there are practical reasons that researchers use samples instead of populations. Perhaps the act of accessing and measuring cases somehow changes or even destroys them. In some types of research, physical materials must be sampled for study because they have to be destroyed in order to be measured. Sometimes researchers must sample because the phenomenon in which they are interested may change in important ways if they take too long to access and measure their cases. For example, relationships between precipitation and snow pack depth must be measured before the season changes. So there are a variety of reasons that sampling is both conceptually and practically necessary in scientific research. But remember that when it is possible and practical to access and measure the entire population of interest instead of taking a sample, it is usually much better to do so. As we make evident below and in Chapter 9, sampling contributes a great deal of complexity and uncertainty to the research process that it would be wonderful to avoid. Unfortunately, such situations are highly rare—the stuff of research fantasy for most scientific geographers.

Sampling Frames and Sampling Designs

Having defined populations as the entire set of entities of interest, and samples as incomplete subsets of populations, we now consider how samples are obtained, and what that means for the design and interpretation of research. We must first recognize that samples are drawn from populations, but that the entire population of interest (sometimes called the "target population") is typically unavailable for sampling. Consider our example of research on urban development. We probably want to generalize beyond cities that currently exist, but of course we can only access and measure cities that do currently exist. The cities we can actually put into our sample are thus a direct subset not of the entire population of interest but of some other subset, a subset that is smaller than our population but larger than our sample. That new subset is called a **sampling frame**—the subset of the population from which cases are actually drawn to become part of the sample. In most research situations, the sampling frame is smaller than the target population, although unlike samples, sampling frames may be the same size as the target population on which they are based, namely, in situations where any member of the target population can potentially become part of the sample. Figure 8.1 shows the conceptual relationship among populations, sampling frames, and samples.

What constitutes a researcher's sampling frame depends on the way he or she identifies and accesses potential cases for measurement. For example, a researcher might identify people from a list of registered voters; he or she might solicit their participation by calling them on the phone. The sampling frame is all of the people on the list of registered voters whose phone number is correct and who are in town during the study period in order to receive your phone call. As another example, a researcher might identify stands of a particular species of cactus by driving along a dirt service road through a mountain range. The sampling frame in this example is all of the cacti visible from all of the roads that the researcher might choose to drive along. Geographers identify cases from property or tax records, industrial records, phone books, census data, maps and images, students in particular classes, people standing on certain street corners, areas that they can walk or drive or sail or fly to,

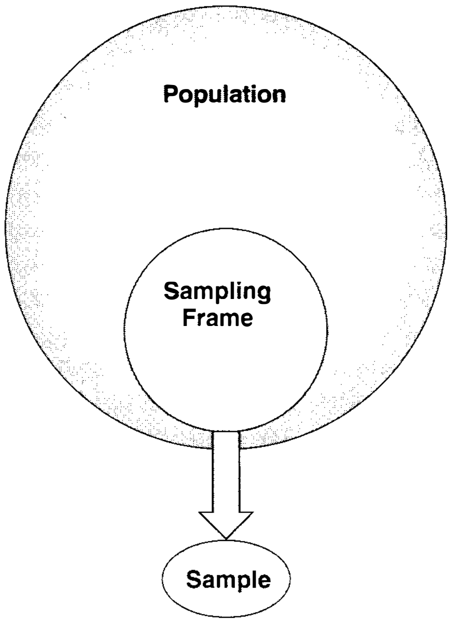

Figure 8.1 The conceptual relationship among population, sampling frame, and sample.

and so on. They solicit the participation of the cases (if the cases are entities with rights, such as individual people or companies) through face-to-face contact, on the phone, through the mail, over the Web, and more.

Once a sampling frame is obtained, the question arises as to how to select the actual sample from the frame. The procedure used to identify cases from the sampling frame to go into the sample is called the **sampling design**. There are several different sampling designs, each of which has different implications for the statistical relationship between the sample and the sampling frame. That is, sampling designs determine what we can know ahead of time about the chances that a case in the frame can get into the sample (by definition, every case in the frame has a nonzero chance of being sampled). These different designs can be grouped into two main categories: **nonprobability sampling** and **probability sampling.** In nonprobability designs, the probability of a particular case being selected is unknown—the researcher using a nonprobability design cannot say beforehand what the precise chance is of any particular case being selected to become part of the sample. Specific nonprobability designs include **convenience sampling**, in which the researcher simply accepts every case he or she can conveniently get hold of to be in the sample, up to the point where the sample is large enough. Obviously there are many specific ways this can be carried out; an example would be geographers who accept all of the remotely sensed imagery for an area they can get hold of, at whatever time and resolution it happened to be recorded. Mining pits provide a convenient opportunity to study the stratigraphy of sedimentary rock; however, it's unlikely that the location of the mining pit was chosen in order to provide the most informative data to a researcher. Another specific nonprobability design is **snowball sampling**. This is a design in which the researcher uses a case he or she has already put into the sample to find out about further cases that could be selected. A researcher studying the spatial behavior of heroin users might ask one user about other users.

In contrast, probability sampling includes any design in which the probability of each case in the sampling frame has a known probability of being selected. There are a variety of specific probability sampling designs used by geographers. The simplest and most common is **simple random sampling**. In this design, each member of the sampling frame has an equal chance of being selected into the sample. In addition, however, each equal-sized *subset* of members of the sampling frame has an equal chance of being selected into the sample (Box 8.1 discusses ways to generate random numbers for sampling and other uses). That is, selecting any particular member of the sampling frame has no effect on the probability that any other particular member will be selected. This second characteristic distinguishes simple random sampling from **systematic random sampling**, in which a first member is randomly selected, then every "nth" member after that is selected. For example, a county could be selected randomly from a list; then every 10th county after that could be selected. In such a sampling design, each county has an equal chance of being selected, because each has an equal chance of being selected first. But each subset of counties does not have an equal chance of being selected—two counties next to each other on the list cannot both be in the sample, for instance. Systematic sampling potentially poses a problem if there is some nonrandom pattern within the sampling frame list. Alphabetic ordering may contain such patterns, as when close places are similarly named. If the list is itself randomly ordered, systematic random sampling is equivalent to simple random sampling.

There are more complex probability sampling designs besides simple and systematic random sampling. In **stratified random sampling**, the sampling frame is segmented into subsets or classes called "strata," based on some relevant variable. Potentially useful stratification variables in geography could include sex, age, race or ethnicity, political party, socioeconomic status, species, soil type, season, ecosystem, altitude, and so on. Even literal "strata" in sedimentary rock could provide a basis for stratified sampling (and a decent pun, no doubt). Of course, stratification is only possible if you know the class of each potential case on the stratification variable. Assuming you do, you would then randomly select the number of cases from each stratification class that matches its proportionate size in the sampling frame.[1] If 7.8% of the households in the sampling frame include four or more children, for example, then cases would be selected to make sure the sample

[1] In some research studies, cases are selected equally from classes rather than proportionately, so that the number of cases from each class is equal. When this is done, the resulting sample will not represent the sampling frame very closely, but the statistical power of comparisons between stratification classes will be maximized. This sampling design makes sense if one is more interested in comparing stratification classes than in characterizing the population as a whole.

Box 8.1 Generating Random Sequences for Sampling and Other Uses

Researchers often need to create random sequences. They are required for simple random sampling and other forms of probability sampling. They are also needed to create random assignments of cases to conditions in experimental studies (Chapter 7). Stochastic numerical models (also Chapter 7), such as those used in transportation or climate simulations, usually include a simulation of random processes. How does one create these random sequences? If one needs only a few random digits, any simple physical system that generates random patterns can be used, including rolling dice, drawing shuffled cards, or flipping coins (although recent statistical research suggests the original upward face has a slightly greater chance of coming up after the toss!). To use these simple methods, assign a digit to each potential entity, making sure that the system you use generates enough different digits to cover all of your alternatives (for example, a single roll of the die produces only six possible outcomes, insufficient for ordering seven entities). If you can use the same entity more than once in a given sequence, called **selecting with replacement**, you can accept sequences in which the same digit comes up more than once; in fact, you must use such sequences that include repeating digits because you would be violating randomness otherwise. If you need sequences in which each entity is selected once and only once, called **selecting without replacement**, you should exclude a digit that repeats one you have already selected in that sequence and go to the next one.

When you require larger numbers of digits, you can use tables of random numbers (found in many statistics books). Here is a small example of such a table that we generated:

47359872271185676654332724051291678701662806568797288431 0914

29363528841075314372317841658558082282704060388836520889887 5

04302510310979592261854670839466912768993746203190959507931 3

02301823752371420223402401980777367395242374354632119643624 9

83666157955576100002100657826552475373840283402624564713815 1

When using random number tables, you can start anywhere and move in any direction as you go along; in fact, it is best to start in different places and move in different directions when using the table on different occasions. Assign digits to entities appropriately. If you have seven entities to randomize, you could use the digits 0–6, skipping any occurrences of the digits 7–9. If you have more than 10 entities to randomize, treat the digits as double digits. Given 30 entities to randomize, you could use the double digits 00–29, skipping any occurrences of the digits 30–99. Make sure that you make all choices with digit sequences of the same length. That is, don't choose some entities with a single digit and others with double digits—they are not equally probable.

Of course, in many situations, it is more convenient to use a computer program to generate the random sequences (technically they are "pseudo-random"); tables of random digits are created this way. Geographers who practice stochastic computational modeling necessarily generate random digits this way. Many programs can create random digits with various properties, including statistical, mathematical, and even spreadsheet programs. As in the less "technological" methods described above, random number programs can be programmed to sample with or without replacement, and with many other restrictions that might be appropriate in particular sampling situations. For example, researchers doing telephone surveys often use some form of "random-digit dialing" to obtain probability samples of respondents. In that situation, it is a good idea to generate only the telephone suffixes (the last four digits) randomly, restricting the area codes and prefixes to those actually in use in the region being sampled; a great many phone numbers generated entirely randomly will not be in use in a region.

contained, as nearly as possible, 7.8% households with four or more children. Compared to simple random sampling, stratified random sampling reduces sampling error (the variability of random samples from one another—see Chapter 9) because all possible samples will be the same in terms of the number of cases they contain from each stratification class. Importantly, stratification does not reduce sampling bias *on average* as compared to simple random sampling, but it does guarantee that any single sample is unbiased on the stratification variable. Stratified sampling is thus useful to ensure that small or important classes are adequately sampled. These benefits sound rather impressive, but it is important to recognize that they are only *potential* benefits. They are realized to the degree that the strata are internally homogeneous but different from the other strata. And the stratification variable must be something statistically related to the other variables that are measured as part of the research. If all ethnic groups have about the same average attitudes about oil drilling, for example, stratifying by ethnicity in a study of attitudes about energy sources will make no difference to your data analysis, although it will allow you to assure consumers of your research that you have represented all ethnic groups proportionately.

Geographers sometimes use **cluster sampling**. The geographer first chooses one or two geographic areas or features, such as cities, parks, or neighborhoods, typically out of convenience rather than randomly. Then he or she randomly selects clusters of cases, such as individual businesses or vernal pools, within each area. This saves costs, as compared to simple random sampling, but it can readily introduce bias in the sample because many characteristics are not evenly or randomly spaced out over areas (as any geographer knows!), so that cases within a single area can be rather different than those in another area. The severity of this bias is reduced if more than just one or two different areas are sampled.

Another probability sampling design often used in geographic research is **multistage area sampling**. The spatial extent of the sampling frame is divided into geographic areas or features, as in cluster sampling. A certain number of these areas, more than one or two, are randomly selected. Then each of these areas is divided into smaller areas or features; a certain number of these smaller areas are randomly selected within each of the larger areas. This can be repeated as much as needed to get down to the desired unit of analysis. For example, five U.S. states may first be selected, then five counties within each state (Louisiana and Alaska would be excluded because they don't have counties), then five census tracts within each county. Multistage area sampling is essentially spatially hierarchical stratified sampling. It has potential benefits similar to those of stratified sampling, but based on geographic areas rather than thematic stratification variables. It is an attractive sampling design from the perspective of a geographer, because it samples from different spatial areas with a spectrum of human and natural characteristics, variations across space that are of such central importance and interest to geographers. Unfortunately, the spatial areas that data come attached to may not match the spatial areas in which structures or processes of interest actually occur. The choice of the appropriate geographic units, in terms of scale and location, is an important and substantial question that we consider in some detail below and in Chapter 9.

Implications of Sampling Frames and Designs

It is important to remember that the cases that provide our data make up the sample, whereas the cases to which we ultimately want to generalize make up the population. But as Figure 8.1 suggests, the relationship of the sample to the population depends both on the relationship of the sample to the sampling frame and that of the sampling frame to the population. So it is important to consider both relationships when designing or interpreting research. For both relationships, the essential issues to consider are the same. The first issue concerns how representative the smaller set (sampling frame or sample) is of the larger set (population or sampling frame, respectively). **Representativeness** is the degree to which the smaller set resembles the larger set. A second issue concerns the question of which is the specific larger set (population or sampling frame) to which we should generalize from the smaller set (sampling frame or sample, respectively). This is the question of **generalizability**—what larger set can we validly draw conclusions about from the evidence of the smaller set? Our choice of a sampling frame determines what population our sampling frame is representative of, and to which population it generalizes. Likewise, our choice of a sampling design determines what sampling frame our sample is representative of, and to which sampling frame it generalizes.[2]

Clearly, the use of a nonprobability sampling design means the researcher cannot say with much certainty how well the sample represents the sampling frame, which means the link from the sample to the population is more uncertain.[3] Nonetheless, nonprobability sampling designs are actually quite common in

[2]Sampling frames and designs have additional implications for increasing or reducing variation due to sampling error in our data, as we discussed in the context of stratification. They can affect the precision of estimation (confidence interval width) and the power of hypothesis testing (chance of finding population effects that actually exist). We also discuss precision and power in Chapter 9.

[3]Even with perfectly conducted probability designs, researchers never know for sure how representative the sample is (except on strata or cluster variables). Such a sampling design only ensures representativeness *on average*, but researchers work with a particular sample, not the average sample. That's why probability designs make the link from sample to sampling frame (and hence to population) *more* certain rather than *perfectly* certain.

geography, and in other natural and social sciences. In some disciplines, and in some topical areas of geography, most sampling of cases is done in a nonprobability manner. Researchers who study rivers rarely sample rivers probabilistically, for example. When one remembers that many entities are sampled in research, not just cases, it becomes even clearer that various forms of nonprobability sampling are actually quite common in research. For example, little or none of even the most carefully conducted survey research asks respondents questions that are probabilistically sampled from some population of all possible ways of wording given questions.

Does this mean that a great deal of research, maybe most of it, is of little worth? In particular, is the common lack of a probabilistic sampling design a critical flaw, as many statistics and methods textbooks claim or imply? Some readers may find it surprising that we think the answer is *no*. In fact, a focus on sampling randomness and representativeness varies greatly across disciplines, subdisciplines, and research purposes. Sampling representativeness is more important for some research goals, such as the estimation of population parameters by pollsters, than others, such as testing hypotheses about causal relationships among constructs. Also, remember that the link from samples to populations always goes through sampling frames, which are typically not identical to either the sample or the population. Sampling designs tell us about the link from samples to sampling frames, but what about the equally important link from sampling frames to populations? In fact, sampling frames are nearly always based on feasibility rather than representativeness. If it were feasible, all researchers would use sampling frames that were as close as possible, if not identical, to the population to which they hope to generalize. Only the difficulty or impossibility of obtaining such a sampling frame causes researchers to use incomplete sampling frames that are often nonrepresentative of the population. For example, geographers who study migration would generally be thrilled to select from a sampling frame that included all people who are moving, who have ever moved, or who will ever move. But this is impossible, and not only because many past and future migrators are deceased or not yet born. Even many migrators currently living don't speak the language of the researcher, or they live on another continent, or they are homeless, or they are inaccessible for some other reason. Instead, such a geographer might accept the use of short-form data from two decades of the U.S. census (Chapter 6), even though that does not come close to reflecting the activities of all American migrators ever, let alone all human migrators ever. Thus, we see that even researchers who go to great lengths to administer probability sampling designs still use sampling frames that may not represent the population of interest all that well. When you realize that the population of interest in basic science so often includes past and future cases, you realize that rarely does a given sampling frame ever allow true probability sampling from that population.

We do not mean to imply that these truths about the way research is actually conducted are not problematic or have no consequences for how research should be interpreted. Either the relative rarity of true probability designs or the common use of only feasible frames, or both, is in fact a nearly ubiquitous shortcoming of research studies. The shortcomings can partially be overcome with more and better-designed research studies, but in most situations, only partially. However, as we have just considered and look at further in Chapter 11, these shortcomings are less important in some research areas than in others and are typically not fatal flaws in research. However, they should always be considered as possible weaknesses of a particular study, and such considerations constitute part of the "Discussion" section of a research paper or talk (Chapter 13).

Nonparticipation and Volunteer Biases

Above, we mentioned the need to solicit the participation of cases like individual people or companies because, of course, these cases have the ethical right to refuse participation in your research (more on research ethics in Chapter 14). There is a specific threat to the representativeness and generalizability of research involving cases like this that arises when some of the cases that are given the opportunity to participate choose not to do so. If nonparticipants are different from participants, the possibility of a **nonparticipation bias** exists. Clearly the sample may become considerably less representative of the sampling frame if the nonparticipation rate is fairly high, but only if nonparticipants differ from participants in ways relevant to your research goals. For example, research on environmental attitudes may be biased if people with "pro-business" views systematically decline to participate because they perceive the researcher and/or the research to have a "pro-environment" slant.

It is always a good idea to record the number and characteristics of nonparticipants, as much as possible. When you do this, you can compare the characteristics of nonparticipants to those of participants or the entire sampling frame. Try to find out why they refused to participate, perhaps during a more intensive follow-up. Of course, you can and should take steps ahead of time to maximize participation, not only because of potential bias but for plain old efficiency.[4] There is a great deal of available wisdom[5] on the varying effectiveness of different forms of contacting potential cases. Personal (live) appeals are nearly always most effective, but of course they are very time consuming. Research has shown that if a person publicly agrees to something, he or she is fairly likely to carry through with it. Thus, survey response rates are higher if potential respondents are first contacted and agree to participate. The use of mail or telephone reminders helps. E-mail requests and reminders to participate can be used nowadays as well, although the rate at which people ignore such requests may be higher even than it is for regular mail. Providing incentives can help too, although researchers usually don't have enough resources to offer really valuable incentives; this obviously depends on the type of potential cases in your sampling frame. Anything that can clarify and shorten a study, such as shortening a survey, can improve participation rates; so can making the study more interesting or relevant sounding to potential participants. If sensitive subject matter is the focus of the

[4]Nonparticipation can lead to the general problem of insufficient sample size, discussed in the next section. In most research situations, this is dealt with simply by making sure to solicit participation from more cases than are needed.

[5]Wisdom about the effectiveness of different types of appeals comes from a great deal of systematic research but also from the experiences of many politicians and sales people. After all, an appeal to participate in a study is essentially a sales pitch.

research, increasing the apparent anonymity of the procedure can help. Working with more carefully trained research assistants that have better "people skills" can obviously increase participation rates too.

On the flip side of nonparticipation bias is the possible sample bias that arises when cases get into the study by volunteering to participate—the so-called **volunteer bias**. This is also known as "self-selection bias" because cases get into studies by selecting themselves. Of course, many studies that involve nonprobability sampling designs are conducted on cases that have, for whatever reason, volunteered to participate. There is quite a bit of research on the characteristics of volunteers as compared to those who tend not to volunteer. For example, volunteers score higher on certain personality dimensions, such as empathy. Most of the methodological issues involving volunteer bias, as well as the solutions, are similar to those for nonparticipation bias.

Spatial Sampling From Continuous Fields

Geographers sample entities and events that are located in space and time, as do all other scientists. Unlike in some other scientific disciplines, however, the spatiality and temporality of cases is often of central importance in geography. Depending on your topical area within geography, you must carefully consider the spatial and/or temporal distribution of your cases when you create your sampling frame and design. In Chapter 5, we presented formal observation schedules for sampling events in time, specifically human behaviors in that chapter. We noted there that time and event sampling are applicable to temporal sampling in all areas of geography. Here we focus on sampling in space. We consider some of the distinctive implications that the spatial distributions of cases have for how we should sample them.

We already touched on spatial sampling in this chapter, in the section above on sampling designs. Both cluster sampling and multistage area sampling are based on obtaining discrete cases distributed in spatial clusters or regions. This is typically the situation in human geography, where cases include such discrete entities as individual people, families, neighborhoods, cities, businesses, schools, governmental institutions, provinces, countries, and so on. In physical geography, in contrast, sampling is very often done from entities such as oceans, rivers, the terrestrial surface, soil layers, and the atmosphere. The properties of these entities that interest geographers are often phenomena that are continuously distributed in space, including many of the physical properties we discussed in Chapter 4: elevation, temperature, precipitation, air pressure, salt concentration, CO_2 concentration, solar energy, and so on. These continuous distributions are known technically as "fields" in contrast to "objects" (we consider the conceptual distinction between fields and objects further in Chapter 12). When studying continuously distributed properties, geographers sample locations within the fields and measure the properties at those locations. Our focus in this section is on this spatial sampling of continuously distributed phenomena.

We noted in Chapter 2 that all measurement must necessarily have finite precision, so that actual data of any type always consist of discrete values. By the same token, we can apparently conceive of properties as being truly continuously distributed in two or three dimensions of space, but our sampling from the space necessarily has finite precision. These considerations point to the strategy geographers use to sample from continuous fields. Organize or break continuous space into discrete objects, perhaps very small and numerous objects, and then sample and measure from these objects. The plans for creating and choosing the discrete objects together constitute the sampling frame and design for sampling from a continuous field distribution. The objects may be points, lines, areas (polygons), or volumes. Of course, the first three objects are not true geometric points, lines, and areas, but rather three-dimensional entities that can be treated for many analytic purposes as if they were zero-, one-, or two-dimensional.

For example, we learned in Chapter 4 that rainfall is measured by rain gauges that can be treated as being at point locations sampled from the field of the two-dimensional land surface; obviously, the openings to rain gauges must have nonzero area or no raindrops at all could fall into them. How many gauges do we need to estimate the continuous field of precipitation well? Where should we place these rain gauges? Or consider taking samples from vegetation assemblages. A common way to do this is to create linear features called **transects**. The word "linear" suggests the approximate one-dimensionality of transects; that is, they are thin or thick "bands." They are not necessarily straight lines, although that often makes sense for sampling because straight lines will be uncorrelated with distributions that are not linearly distributed along the direction of the transect—the straight line can be a convenient way to achieve sampling randomness.[6] Once transects are created along the ground, they can be treated as individual cases by recording all instances of the plants of interest along the transect, or further sampling discretization can occur by recording measurements at sampled points along the transect.

The entire region from which your discrete features could be sampled is your sampling frame.[7] The particular locations you actually sample and measure depend on your sampling design. Just as with nonspatial sampling, discrete feature locations could be chosen with the use of either a nonprobability or probability design. For instance, often enough, locations are chosen by convenience, definitely a nonprobability design. However, if a probability design is used, there are a large variety of them that could be applied. Devising probability sampling in space and place is more elaborate (and more intriguing) than probability sampling in the thematic or temporal domains, because spatiality is two- or three-dimensional instead of

[6]The frequent desirability of straight transects is one reason that physical geographers such as biogeographers are fond of driving along wilderness or mountain service roads to do their sampling and measurements. Erecting power lines is costly in labor and materials, so utility companies go to great lengths to minimize their total length by building straight paths through forests and over rolling topography.

[7]Discretizing fields into self-contained sampling frames is difficult when processes flow into and out of the frame boundaries. The study of "island biogeography" is so important, for instance, because islands bound processes that would not be bounded on larger landmasses.

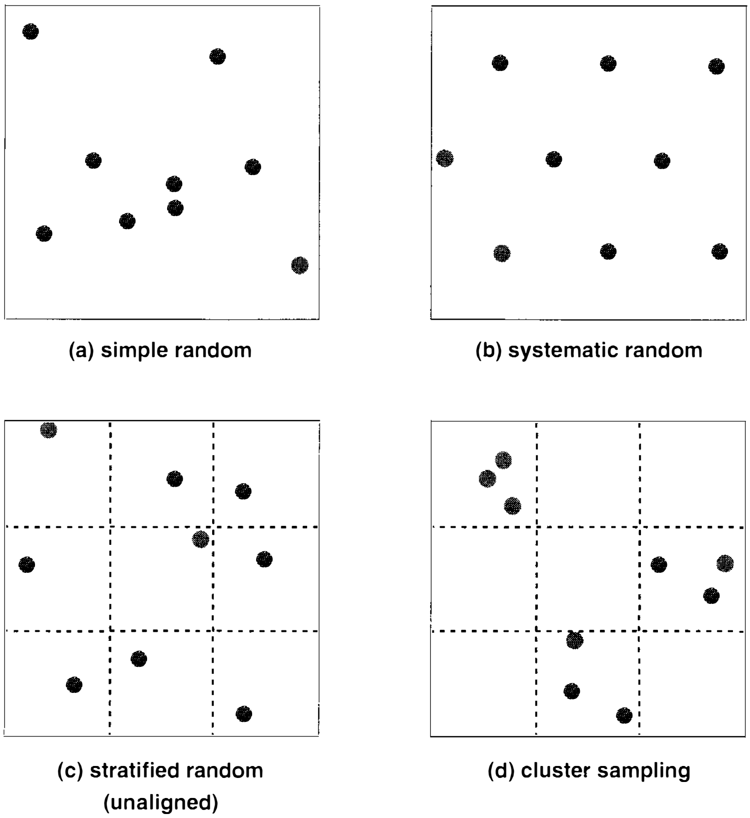

Figure 8.2 Spatial sampling of points from fields.

one-dimensional. Figure 8.2 shows four examples involving point features sampled within a bounded sampling frame. The points could be sampled according to a simple random design. Implement this by imposing a coordinate grid over the total sampling frame, randomly choosing an "X" and a "Y" value for each point you need. You could sample the points according to a systematic random design, where you choose one point at random and then sample the rest at a set distance away to be equally spaced. Sometimes geographic researchers break continuous space into discrete polygonal features shaped like squares—called **quadrats**. These quadrats are analogous to the nonspatial strata we discussed above. Point locations are sampled from within the quadrats. Figure 8.2 shows stratified random sampling, whereby one point is chosen within each quadrat. The location of the points within different quadrats is random, so this is "unaligned" sampling. Finally, cluster sampling can be carried out by sampling quadrats first, then sampling points within each chosen quadrat. This is the same as the cluster sampling discussed above that started with discrete features to begin with.

These approaches to probability sampling all ignore the actual distribution of features or properties of interest; they are examples of **independent spatial sampling**. Think about it for a moment. If you wanted to sample the topography of

South America, its profile of elevation features like mountains and canyons, would it be most efficient to sample and measure elevation from a set of equally spaced points? Would the Amazon Basin require as many points as the Andes Mountains to produce a representative sample of elevations? No, it certainly would not. Often, geographers do not sample from fields as if they were homogeneous. They concentrate their sampling where they believe their property of interest exhibits more variation in reality. That is, geographers sometimes sample on the basis of a model of patterns or **trends** in the spatial distribution of their property of interest; they practice **nonindependent spatial sampling**. In that case, sampling is focused on locations of greater change in the trend. The location and/or orientation of transects, for example, is often done nonrandomly (or at least not fully randomly). It is done in a reasoned way, based on knowledge of the feature distribution being sampled. Thus, transects are placed at right-angle orientations across streams or rivers, up mountainsides rather than along them, and so on.

After they complete sampling and measurement, geographers will typically want to make inferences back to the continuous field. This is the subject of statistical inference from samples to populations that we discuss in detail in Chapter 9. In the context of the *spatial* statistical inference we are concerned with here, this process is called **spatial interpolation**. Spatial interpolation is a simple concept to understand; however, the mathematical techniques used to accomplish it can be rather sophisticated. To appreciate the concept of spatial interpolation, suppose you must provide an estimate of the temperature in Denver, Colorado, if the temperature at Fort Collins (~60 miles north of Denver) is 100°F and the temperature at Colorado Springs (~60 miles south of Denver) is 90°F. If you guess 95°F, you have succeeded at simple spatial interpolation.

We touch on the topic of spatial interpolation again in Chapter 12. Interpolation is a form of educated guessing, so naturally we should be curious as to how accurately it can actually regenerate the continuous field from which samples are taken. The accuracy of spatially interpolating the surface—the spatial distribution of property values—of a field from point measurements depends on[8]

a. the accuracy of measurement

b. the density of sample points (the number of points per area)

c. the spatial distribution of points (largely due to the sampling design)

d. the particular interpolation procedure used (see Chapter 12)

e. the actual spatial distribution of the surface being measured and interpolated.

In order to sample the full detail of a field, it is necessary to sample enough points (to sample at a sufficient density) so that the distance between points is less than the size of any relevant features on the surface you want to capture. According

[8]MacEachren, A. M., & Dsavidson, J. V. (1987). Sampling and isometric mapping of continuous geographic surfaces. *The American Cartographer, 14*, 299–320.

to the "sampling theorem," this distance is one half or less of the length of the smallest feature, whether a "hill" or a "valley" in the values of your property of interest.[9]

Geographers sampling from continuous fields must, therefore, answer questions about how many points should be sampled, the locations from where they should be sampled, and how they should be interpolated. Perhaps more important, however, is the question of whether interpolation should be carried out in the first place. Is the field model really an appropriate conceptualization of your phenomenon? This is the fundamental question of the "ontology" of geographic reality, to which we return in Chapter 12.

Sample Size

Besides issues that concern how samples are obtained, creators and consumers of research must always address the question of *how large* samples of cases should be. The answer to this question is a compromise between two competing motivations—the benefits versus the costs of larger samples. On one hand, larger samples are more likely to be representative of the sampling frame. Larger samples allow a researcher to test more variables and more interaction effects among variables (see Chapter 7), and they allow the researcher to conduct analyses within more subcategories of types of cases. More cases increase the precision of estimation and the power of hypothesis testing (as do particular sampling frames and designs, see footnote 2). Clearly, larger samples bring many benefits to the research enterprise. Given unlimited resources, researchers would almost always want larger samples. But researchers don't have unlimited resources; that's one of the main reasons they sample in the first place. So against the motivation to obtain larger samples is the cost of obtaining them. Larger samples generally cost more money, more time, and more effort. At some point, the additional precision and power gained by increasing sample size is minimal (the marginal benefit becomes tiny), but the increase in cost is not likely to be minimal. Furthermore, it's impossible in some situations to obtain larger samples; for example, a geographer studying community responses to tsunamis will probably have difficulty finding many such communities.

Taking the competing motivations into account, we can give some approximate guidelines on desirable minimum sample size. It is important to recognize that desirable sample size varies greatly with one's research goal, and that different scientific disciplines, as well as subdisciplines of geography, have quite different traditions with respect to sample size. As we discussed in Chapter 7 on research design, quality research is sometimes done with a single case. Early exploratory phases of research often focus on just one or a few cases. In fact, we strongly recommend the wisdom of trying out one's study procedures (instructions, measurement tools, question wordings, stimulus materials, analysis procedures, and so on) in a trial pilot study employing a small sample of cases similar to those you will use in the regular study,

[9]See, for example, Tobler, W. (2000). The development of analytical cartography. *Cartography and Geographic Information Science, 27*, 189–194.

no matter how large a sample you eventually intend to obtain. With some research questions, only one case is logically required to achieve a respectable level of confidence in one's conclusions. For example, if you wanted to show that it is possible for an artificial reef to sustain populations of a particular fish species, you would need only one such reef to establish this. In contrast, large samples of one to two thousand or more are required to estimate the parameters of some population with great precision. This number must be especially large if the makeup of the population is very heterogeneous with respect to your variable of interest.[10] Political pollsters, for instance, must obtain large samples of a few thousand in order to predict national election outcomes accurately (although as a percentage of all voters, this is surprisingly small). Between these two sample-size extremes, researchers who are testing hypotheses about causal relationships—theories—traditionally try to get at least 20–30 cases.[11] This recommendation derives from a tradition of making sure to have enough power, which increases with increasing sample size, to achieve statistically significant results given a weak to moderate relationship in the data (a correlation of approximately .3). If more than two variables are to be analyzed, or more complex comparisons are to be made (for example, nonlinear as well as linear relationships are to be analyzed), this number should be increased.

Formal **power analysis** techniques have been developed, and they can be modified a little to perform **precision analysis** for statistical estimation. A primary purpose of these techniques is to determine how large a sample is required to obtain statistically significant results or a confidence interval of desired width, given assumptions about the size of relationships in the population and the amount of noise in the data. Hopefully, these assumptions are based on careful reasoning, perhaps on prior research. The size of the relationship expressed as a proportion of noise in the data is called an **effect size**. In some situations, one can decide on an effect size that would be meaningful in a particular research context and aim for that effect size. These techniques are also used to evaluate nonsignificant results; as we discuss in Chapter 9, statistical nonsignificance does not prove that there is no relationship in the population. It may be that the effect is too small to be picked up robustly with a given number of cases. These considerations are part of the quantitative technique of reviewing research called "meta-analysis," which we discuss in footnote 1 of Chapter 13.

---

[10]Contrary to most people's initial intuition (including our own), the sample size required to estimate a population parameter precisely does not depend on the size of the population.

[11]This number is based on the simple situation of calculating relationships by correlating two variables measured on each case. Alternatively, in experimental studies, relationships are often measured by comparing the means of one variable across groups or conditions of another variable (these concepts were covered in Chapter 7). In the simplest such experiment, the means of two groups of cases are compared (for example, the readability of two map designs is compared). In this experimental design, 20–30 cases per group are considered a desirable minimum. If the same set of cases can be placed in both of the groups at different times (that is, a within-case design—see Chapter 7), this number can revert back to a total of 20–30 cases or even less.

Review Questions

- What are populations and samples?
- Why do geographic researchers sample? What are some things that geographers sample, and why do they sample them?

Sampling Frames and Sampling Designs

- What are sampling frames and sampling designs, and how do they relate to populations and samples?
- What is the distinction between probability and nonprobability sampling designs, and what are the implications of this distinction for research?
- What are the following types of sampling: simple random, systematic random, stratified random, cluster, multistage area?

Implications of Sampling Frames and Designs

- What is sample representativeness, and how does it influence generalizability?
- How common are nonprobability-sampling designs in geographic research, and what are the implications of using such sampling designs?
- What are nonparticipation and volunteer biases, and what are some approaches to minimizing their negative effects on research?

Spatial Sampling from Continuous Fields

- What are some implications of sampling from sets of discrete objects versus sampling from continuous fields?
- What is the general approach usually taken to sampling from continuous fields, and what are some difficulties that arise from this approach?
- What is spatial interpolation, and what are some factors that influence its accuracy?

Sample Size

- What are the competing motivations for larger and smaller samples?
- What are the techniques of power analysis and precision analysis, and how can they be used to help decide on sample size?

Key Terms

cluster sampling: specific type of probability sampling design in which cases in the sampling frame are grouped into spatial areas or clusters, and then cases are selected randomly from each cluster to be in the sample, usually proportionately in number to the cluster size

convenience sampling: specific type of nonprobability sampling design in which researchers take every case from the sampling frame they can conveniently get hold of to be in the sample, until their sample is large enough

effect size: the size of a statistical relationship expressed as a proportion of noise in the data

generalizability: the validity with which you can draw conclusions about sampling frames from samples, or about populations from sampling frames; depends on representativeness

independent spatial sampling: spatial sampling design that ignores the actual distribution of features or properties being sampled

multistage area sampling: specific type of probability sampling design in which cases in the sampling frame are grouped into spatial areas or clusters (like cluster sampling), smaller areas are defined and randomly selected within the larger areas, possibly more than once at increasingly smaller scales of area, and cases to be in the sample are finally selected randomly from each of the smallest areas

nonindependent spatial sampling: spatial sampling design that is based on the actual distribution of features or properties being sampled

nonparticipation bias: the degree to which a sample is not representative of a sampling frame because potential cases who refuse to participate in the study are different than those who do

nonprobability sampling: any sampling design in which the probability of a particular case being selected from the sampling frame is unknown ahead of time

population: the entire set of entities of interest, including cases, measures, settings, and so on; sometimes called a "target" population

power analysis: formal technique to estimate the statistical power in a particular hypothesis test; primarily used to estimate the sample size required to achieve statistical significance, given a population effect and error variance of particular sizes

precision analysis: formal technique, similar to power analysis, to estimate the width of the confidence interval in a particular statistical estimation; primarily used to estimate the sample size required to achieve an interval of a given width

probability sampling: any sampling design in which the probability of a particular case being selected from the sampling frame is known ahead of time

quadrats: rectangular polygonal features created to sample from continuous field distributions

representativeness: the degree to which the make-up of the sample resembles that of the sampling frame, or the make-up of the sampling frame resembles that of the population; determines generalizability

sample: incomplete subset of the entire set of entities of interest, including cases, measures, settings, and so on; necessarily smaller than the population from which it is drawn

sampling: any way of selecting a sample of entities from a population

sampling design: the procedure used to identify cases from the sampling frame to go into the sample

sampling frame: the subset of the population from which cases are actually drawn to become part of the sample; usually smaller than the target population

selecting with replacement: choosing a subset of outcomes or entities from a larger set in such a way that the same outcome or entity can be chosen more than once; the entity or outcome is essentially "replaced" after it is chosen so it can be chosen again

selecting without replacement: choosing a subset of outcomes or entities from a larger set in such a way that the same outcome or entity cannot be chosen more than once; the entity or outcome is not "replaced" after it is chosen

simple random sampling: specific type of probability sampling design in which each case in the sampling frame, and each subset of cases, has an equal chance of getting selected into the sample

snowball sampling: specific type of nonprobability sampling design in which researchers use cases they have already put into the sample to find out about further cases that could be selected

spatial interpolation: mathematical technique in which unknown data values at locations within a field are inferred from values that are actually measured; essentially spatial statistical inference from a sample of measurements to a population field

stratified random sampling: specific type of probability sampling design in which cases in the sampling frame are grouped into thematic classes or "strata," and then cases are selected randomly from each class to be in the sample, usually proportionately in number to the class size

systematic random sampling: specific type of probability sampling design in which a first member is randomly selected from the sampling frame to be in the sample, and then every "nth" member after that is selected; as compared to simple random sampling, each case in the sampling frame has an equal chance of getting selected, but each subset of cases does not

transects: linear features created to sample from continuous field distributions

trend: pattern in the spatial distribution of properties of interest that strongly influences the adequacy of different spatial sampling frames and designs

volunteer bias: the degree to which a sample is not representative of a sampling frame because potential cases who volunteer to participate in the study are different than those who do not; also known as self-selection bias

Bibliography

Fowler, F. J. (2002). *Survey research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Haining, R. (1990). *Spatial data analysis in the social and environmental sciences.* Cambridge: Cambridge University Press.

Murphy, K. R., & Myors, B. (2004). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Webster, R., & Oliver, M. A. (1990). *Statistical methods in soil and land resource survey.* Oxford: Oxford University Press.

Weisberg, H. F., Krosnick, J. A., & Bowen, B. D. (1996). *An introduction to survey research, polling, and data analysis* (3rd ed.). Thousand Oaks, CA: Sage.