INTRODUCTION

Informant accuracy, data validity, and even ethical issues (like whether it's all right to deceive people in conducting experiments) are all measurement problems in research. The other big class of problems involves sampling: Given that your measurements are credible, how much of the world do they represent? How far can you generalize the results of your research?

The answer depends, first of all, on the kind of data in which you're interested. There are two kinds of data of interest to social scientists: individual data and cultural data. These two kinds require different approaches to sampling.

WHAT ARE SAMPLES AND WHY DO WE NEED THEM?

Individual data are about **attributes of individuals in a population.** Each person has an age, for example; each person has an income; and each person has preferences for things like products, political positions, and characteristics for a mate. If the idea in collecting data is to estimate the average age or income or preference in a larger population—that is, to estimate some ***population parameters***—then a scientifically drawn, unbiased sample is a must. By "scientifically drawn," I mean random selection of cases so that every unit of analysis in your study has an equal chance of being chosen for study. Barring that (since it's often impossible to choose a true random sample), I mean a sample where the probability of inclusion is known for each case.

Cultural data are different. Cultural data require experts. If you want to understand a process—like how people in a factory work group decide on whether to lodge a complaint to management, or how police in a squad car determine whether to stop someone on the street—then **you want people who can offer expert explanations and who represent the intracultural variation that we find in all societies.** It's one thing to ask: "How many people did you stop on the street for questioning last week?" This requires an answer about individual behavior. It's another thing to ask: "How do people in your squad decide whether to stop someone for questioning on a street patrol?" This requires cultural experts.

All social scientists are interested in both kinds of data. An anthropologist may ask people: "How many cows did *you* give to your in-laws as bride price when you got married?" They may also ask key informants (experts), "So, why do men who get married around here deliver cows to their in-laws?"

Individual-attribute data require *probability sampling*; cultural data require *nonprobability sampling*. Probability sampling is by far the best known kind of sampling in the social sciences, so we'll begin with it. This will take us into a discussion of probability theory, variance, and distributions. As we go through that material, remember that Freud's ideas about the unconscious, Skinner's ideas about operant conditioning, and Piaget's ideas about stages of cognitive development were all based on intensive study of a relatively small number of cases and not on a survey of a random sample of anything.

This is not an argument against random sampling. Far from it. It's just a reminder that there are lots of ways to conduct scientific research. When the project calls for random sampling, it pays to get that right, too. Whether the population consists of 16,886 households in a town, the 4,643 records of residential property transactions between 1970 and 1999 in that same town, or the 40,000 members of the National Cattlemen's Beef Association, it takes less time and less money to study a sample of them than it does to study all of them.

If samples were just easier and cheaper to study but failed to produce useful data, there wouldn't be much to say for them. A study based on a random sample, however, is often better than one based on the whole population.

WHY THE U.S. STILL HAS A CENSUS

Estimates of many segments of the population of the U.S., for example, would be more accurate if they were based on good samples instead of on a census. The General Accounting Office estimated that, in 1990, the Bureau of the Census failed to count 5.3 million people, mostly African Americans and Hispanics (*New York Times,* August 25, 1991, sec. 1, p. 23). This means, among other things, that statistics about the incidence of everything from criminal behavior to AIDS among African American men are skewed.

Lots of things can go wrong with counting. Heads of households are responsible for filling out and returning the census forms, but only 63% of the mailed forms were returned in 1990 (that's down, by the way, from 78% in 1970). The Bureau of the Census had to hire and train half a million people to track down all the people who had not been enumerated in the mailed-back forms.

Some people wound up being counted twice. Some college students were counted twice, for example, because their parents had counted them on the mailed-back census form. Then, on census day, some of those same students were tracked down again by enumerators who canvassed the dorms. Meanwhile, lots of other people (like illegal immigrants and people living in places to which the census takers would rather not go) were not being counted at all.

The Bureau of the Census publishes adjustments to the census figures based on samples, and every so often there is serious talk of replacing the census with estimates based on samples. The U.S. Constitution, however, in Article I, Section 2, *requires* that the government conduct an "Enumeration" (with a capital E) of the population every 10 years to apportion seats in the House of Representatives to the states.

There's also a serious political issue: If sampling produced higher estimates of the number of citizens who are, say, homeless or who are migrant farm workers, this would benefit only certain states. It would also benefit the Democratic Party over the Republican Party, since more poor people vote the Democratic ticket. In 1999, the Supreme Court of the United States held that, at least for the big ***decennial census,*** the word "Enumeration" means going out and counting noses. So, for the moment, we're stuck with a census and all its inaccuracies.

UDDRAG AF: BERNARD, H. RUSELL (2000). SOCIAL RESEARCH METHODS – QUALITATIVE AND QUANTITATIVE APPROACHES

IT PAYS TO TAKE SAMPLES AND TO STICK WITH THEM

If you are doing all the work yourself, it's next to impossible to interview more than a few hundred people. Even a small, county school system might have 500 employees, including teachers, administrators, and staff. You'd need several interviewers to reach all those people within a reasonable amount of time. Interviewers may not use the same wording of questions; they may not probe equally well on subjects that require sensitive interviewing; they may not be equally careful in recording data on field instruments and in coding data for analysis. And, as you'll see in the section on telephone interviewing, in Chapter 7, some interviewers actually falsify data.

Most important, you have no idea how much error is introduced by these problems. A well-chosen sample, interviewed by people who have similarly high skills in getting data, has a known chance of being incorrect on any variable. (Careful, though: If you have a project that requires multiple interviewers and you try to skimp on personnel, you run a big risk. Overworked or poorly trained interviewers will cut corners; see Chapter 7.) Furthermore, studying an entire population may pose a history threat to the internal validity of your data. If you *don't* add interviewers it may take you so long to complete your research that events intervene that make it impossible to interpret your data.

You're interested in how the nursing staff at a midsize, private hospital feels about a reorganization plan. You decide to survey all 210 nurses on the staff, using a structured, 10-minute personal interview. You know that it's tough to track some nurses down—they are very busy and sometimes don't have even 10 minutes to stop and chat; they change shifts, forcing you to find them at four in the morning—but you have three months for the research and you figure you can do the survey a little at a time.

Two months into your work, you've gotten 160 interviews on the topic—only 50 to go. Just about that time, the hospital announces that it has been bought out by a big health maintenance corporation—one that's traded on the New York Stock Exchange. All of a sudden the picture changes. Your "sample" of 160 is biased toward those people whom it was easy to find, and you have no idea what *that* means. And even if you could now get those remaining 50 respondents, their opinions may have been radically changed by the new circumstances. The opinions of the 160 respondents who already talked to you may have also changed.

Now you're really stuck. You can't simply throw together the 50 and the 160 interviews; you have no idea what that will do to your results. Nor can you compare the 160 and the 50 as representing the nursing staff's attitudes before and after the buyout. Neither sample is unbiased with regard to what you are studying.

If you had taken a *random sample* of 60 people in a single week early in your project, you'd now be in much better shape, because you'd know the potential sampling error in your study. If historical circumstances (the surprise buyout, for example) require it, you could interview the same sample of 60 again (in what is known as a *panel study*), or take another representative sample of the same size and see what differences there are before and after the critical event. In either case, you are better off with the sample than with the whole population.

By the way, there is no guarantee that a week is quick enough to avoid the problem described here. It's just less likely to be a problem. Less likely is better than more likely.

WHAT KINDS OF SAMPLES ARE THERE?

There are eight major kinds of samples. Four of them—simple random, systematic random, stratified random, and cluster samples—are based on the principles of probability theory. The other four—quota, purposive/judgment, convenience/haphazard, and snowball samples—are not. *Probability samples* are representative of larger populations and they increase external validity in any study.

The general rule is this: If your objective is to generalize about individual characteristics from a sample to a population, then use probability sampling whenever you can; use *nonprobability sampling* methods when you really, really can't do probability sampling. There are some research problems that simply demand nonprobability sampling. If your objective is to generalize about cultural data (rather than about individual characteristics), then use one of the appropriate methods for nonprobability sampling. More about them later.

PROBABILITY SAMPLING

Probability samples are based on taking a given number of units of analysis from a list, called a *sampling frame*, which represents some *population* under study. In a probability sample, or *unbiased sample*, **each individual has exactly the same chance as every other individual of being selected.**

In 1970, while the United States was engaged in a very unpopular war in Vietnam, men were selected to serve in the military by a supposedly random draw. Three hundred sixty-six capsules (one for each day of the year, including leap year) were put in a drum and the drum was turned to mix the capsules. Then dates were pulled from the drum, one at a time. All the men whose birthdays fell on the days that were selected were drafted.

When enough men had been selected to fill the year's quota, the lottery stopped. Men whose birthdays hadn't been pulled were safe until the following year when the lottery would be run again. It turned out that men whose birthdays were in the later months had a better chance of being drafted than men whose birthdays were earlier in the year. This happened because the drum wasn't rotated enough to thoroughly mix the capsules (Williams 1978). Not a good sampling technique.

Sampling Frames

If you can get it, the first thing you need for a good sample is a good *sampling frame*. (I say "if you can get it" because a lot of social research is done on populations for which no sampling frame exists. More on this later.) A sampling frame is a list of units of analysis *from which* you take a sample and *to which* you generalize.

A sampling frame may be a telephone directory, the tax rolls of a community, or a census of a community that you do yourself. In the U.S., the city directories (published by R. L. Polk and Company) are often adequate sampling frames. The directories are available for many small towns at the local library or chamber of commerce. Professional survey researchers in the U.S. often purchase samples from firms that keep up-to-date databases just for this purpose.

Then we apply the finite population correction:

$$n' = \frac{385}{1 + (384/540)} = 225$$

This is still a hefty percentage of the 540 people in the population, but it's a lot smaller than the 384 called for by the standard formula (see Box 5.5).

Another Catch

All of this discussion has been about estimating single parameters, whether proportions or means. You will often want to measure the interaction among several variables at once. Suppose you study a population of wealthy, middle-class, and poor people. That's 3 kinds of people. Now add two sexes, male and female (that makes 6 kinds of people) and two colors, Black and White (that makes 12 kinds). If you want to know how all those independent variables combine to predict, say, average number of children desired, the sampling strategy gets more complicated.

Sampling is one of the trickiest parts of social research. I recommend strongly that you consult an expert in sampling if you are going to do complex tests on your data. (For excellent coverage of all the basics in sampling theory and sample design, see Sudman [1976] and Jaeger [1984].)

Box 5.5
Settling for Bigger Confidence Intervals

If we were willing to settle for a 10% confidence interval, we'd need only 82 people in this example, but the trade-off would be substantial. If 65 out of 225, or 29%, reported that they had participated in one or more extreme sports, we would be 68% confident that from 24% to 34% really did, and 95% confident that 19% to 39% did. But if 24 out of 82 (the same 29%) reported having participated in extreme sports, we'd be 68% confident that the true figure was between 19% and 39%, and 95% confident that it was between 9% and 49%. With a spread like that, you wouldn't want to bet much on the sample statistic of 29%.

Anthropologists, epidemiologists, and others who work with small populations often face this problem. Anthropologists work around it by relying on long-term ethnography—combining qualitative and quantitative approaches. Epidemiologists may use special statistical treatments appropriate to small samples.

NONPROBABILITY SAMPLING

If your objective is to estimate a parameter or a proportion from a sample to a larger population, and if your research calls for the collection of data about attributes of individuals (whether those individuals are people or Boy Scout troops or episodes of a sitcom), then the rule is simple: Collect data from an unbiased sample. If you know that you *ought* to use a random, unbiased sample and you have the means to *get* an unbiased sample, and you still choose to use a nonprobability sample, then expect to take a lot of flak.

There are, however, two quite different circumstances under which nonprobability samples are exactly what are called for.

(1) Despite our best efforts, it is often impossible to do probability sampling under real research conditions. In these cases, use a nonprobability sample and *document the bias.* That's all there is to it. No need to agonize about it.

(2) As I mentioned at the beginning of this chapter, when you are collecting cultural data, as contrasted with data about individual experience, then expert informants, not randomly selected respondents, are what you need. (See Box 6.1 for more on this distinction. And for a detailed discussion of on selecting expert informants, see Chapter 9.)

There are several nonprobability sampling methods that are appropriate under different circumstances. These include ***quota sampling, purposive sampling*** (also called ***judgment sampling***), ***convenience sampling*** (also called ***haphazard sampling***), and ***snowball sampling.***

Quota Sampling

In quota sampling, you decide on the subpopulations of interest and on the proportions of those subpopulations in the final sample. If you are going to take a sample of 400 full-time employed adults in a city, you might decide that, since gender is of interest to you as an independent variable, and since women make up about half the population, then half your sample should be women and half should be men. Perhaps half of each gender quota should be older than 40 and half should be younger; maybe half of each of those quotas should be self-employed and half should be salaried. How you design a quota sample depends, like any sampling problem, on what's important for your study.

When you are all through designing a multistage quota sample, you go out and fill the quotas. You look for, say, five self-employed females who are over 40 years of age and who are self-employed and for five salaried males who are under 40. And so on. Quota sampling resembles stratified probability sampling with an important difference: Respondents are not chosen randomly. Instead, interviewers choose members of the sample on the spot.

Commercial polling companies use quota samples that are fine-tuned on the basis of decades of research. Organizations like Gallup, Roper, and Harris have learned how to train interviewers not to choose respondents who are pretty much like themselves; not to select only people whom they would enjoy interviewing; not to avoid people whom they would find obnoxious or hostile; not to avoid people who are hard to contact (busy people who are hardly ever home, or people who work nights and sleep days); and not to favor people who are eager to be interviewed.

The result is quota samples that are *not unbiased* but that often do a good job of reflecting the population parameters of interest. In other words, quota sampling is an art that often approximates the results of probability sampling at less cost and less hassle than strict probability sampling.

Often, but not always. Back in 1948, pollsters predicted, on the basis of quota sampling, that Thomas Dewey would beat Harry Truman in the U.S. presidential election. The *Chicago Tribune* was so confident of those predictions that it printed an edition announcing Dewey's victory—while the votes were being counted that would make Truman president.

Half a century later, quota sampling is just as dangerous. In the general election in Britain in 1992, four different polls published on the day of the election put the Liberal Party, on average, about 1 point ahead of the Conservative Party. All the polls were based on quota sampling. The Conservatives won by 8 points. In fact, from 1992 to 1997, political polls using quota samples in Britain systematically overestimated the support for the Liberals (Curtice and Sparrow 1997).

Quota sampling is also used in the study of cultural data, but here the operation and the rationale are entirely different. Anthropologists often speak of choosing "key informants"—people who are knowledgeable about particular domains of life in a culture—and of focusing their attention on those people. We can extend the idea of *a* key informant, to a *set* of key informants—people who are chosen in a quota sample to represent the variation in domains of a culture.

For example, if you want to know about how Little League baseball functions in the life of American families, you wouldn't ask people who have never played the game or who have never had children who played the game. But there is plenty of intracultural variation even among Little League players and parents of players. Open-ended interviews with four or five really knowledgeable people will produce the relevant cultural data. For data on how individuals feel about the level of competitiveness in Little League, or on the prevailing opinion with regard to Little League as preparation for later life, 30 or 40 open-ended interviews will produce the appropriate data.

Purposive or Judgment Sampling

In judgment sampling, you decide the purpose you want informants (or communities) to serve, and you go out to find some. This is somewhat like quota sampling, except that there is no overall sampling design that tells you how many of each type of informant you need for a study.

Kimberly Mahaffy (1996) was interested in how lesbian Christians deal with the cognitive dissonance that comes from being rejected by mainstream Christian churches. Mahaffy sent letters to gay Christian organizations, asking them to put an ad for potential respondents in their newsletters. She sent flyers to women's bookstores and to lesbian support groups, asking for potential respondents to get in touch with her.

Eventually, Mahaffy got 163 completed questionnaires from women who fit the criteria she had established for her research, including 44 from women who self-identified as born-again or evangelical Christians. Mahaffy could not possibly have gotten an unbiased sample of lesbian Christians, but the corpus of data that she collected from her respondents had all the information she needed to answer her research questions.

Moreover, in judgment sampling it's not even necessary to decide up front what *kinds* of units of analysis to study. I used purposive sampling in my study of the Kalymnian (Greek) sponge-fishing industry (Bernard 1987). I knew I had to interview sponge merchants, boat owners, and divers, but my first interviews taught me that I had to interview people whom I had never considered: men who used to be divers but who had quit, gone to Australia as labor migrants, and returned to their island.

There are many good reasons for using purposive samples. They are used widely in *pilot studies* before testing a hypothesis with a representative sample. Pound et al. (1993) developed a questionnaire to measure the satisfaction of caregivers of people who had suffered a stroke. They did the pilot test of the questionnaire on 23 people—caregivers of patients whose names were taken straight out of hospital registration lists. Samples don't get much more purposive than this, and they don't get much more appropriate, either.

Purposive samples are also used in the selection of a few cases for intensive study. Researchers don't usually pull research sites—communities, hospitals, school systems—out of a hat. They rely on their judgment to find one that reflects the things they are interested in. Life history research and qualitative research on special populations (drug addicts, trial lawyers, shamans) rely on judgment sampling.

Hays (1984) went through the obituary section of a local newspaper and interviewed a purposive sample of survivors. She found that three-fourths of those who died at age 65 or older had surviving children or siblings living close enough to provide support and one-fourth didn't. Barroso (1997) studied a purposive sample of 14 men and 6 women in the Tampa, Florida, area, all of whom had lived with AIDS for at least three years.

Judgment sampling is also used in studying critical cases. Polling companies try to identify communities across the United States that have voted for the winner in the past, say, six presidential elections. Then they poll those few communities that meet the criterion. Choosing key informants in ethnographic research is a kind of critical-case sampling. It would be pointless to select a handful of people randomly from a population and try to turn them into trusted key informants (for more on ethnography, see Chapter 9).

Purposive sampling can be applied to any units of analysis, just like probability sampling. In the mid-1960s, in a movement known as "mainstreaming," publicly supported psychiatric hospitals across the U.S. were closed and their residents were released to live in their own communities. A private board-and-care industry has grown up to accommodate these patients, as well as people who suffered from physical disorders. Shostack and Campagna (1991) were interested in the conditions and needs of these board-and-care homes in New Jersey, and in the needs of the residents of those homes—things like whether residents had opportunities for recreation, whether residents were being prepared for life outside the home, and so on.

Working alone and without financial support, Shostack and Campagna were not able to cover the entire state. They approached the welfare boards of five New Jersey counties, and those boards arranged for interviews with the operators of 17 homes in their jurisdictions. Table 5.7 compares the age, sex, and racial distribution of the residents in Shostack and Campagna's purposive sample with that of the 450 homes in the entire state.

This is an interesting example of nonprobability cluster sampling. Shostack and Campagna identified appropriate board-and-care *homes* and then interviewed the operators of those homes. Shostack and Campagna are careful to note that theirs is a judgment sample, not an unbiased sample. They had a very good sample for an

TABLE 5.7 Characteristics of Residents in Shostack and Campagna's (1991) Sample Homes and All of the Homes in the State of New Jersey in 1988

Characteristics	*Residents in Sample Homes (percentage)*	*Residents in All 450 Homes (percentage)*
Male	43	42
Female	57	58
Under 50 years of age	19	31
50 and over	81	69
White	89	88
Non-White	11	12

SOURCE: Adapted from A. E. Shostack and G. P. Campagna, *Adult Residential Care Journal* 5. Copyright © 1991. Reprinted by permission of Plenum Publishing Corporation.

exploratory study, however, especially one that involved just two researchers. Their detailed findings about the institutional needs of the low-income frail elderly and the low-income mentally ill who live in privately run board-and-care homes deserve careful reading.

Most purposive samples are pretty small (ethnographic studies are usually based on 30-60 interviews, plus participant observation), but purposive sampling is not just for small samples. Ackerman and Gondolf (1991) had some ideas they wanted to test about differences between ACOAs (adult children of alcoholics) and non-ACOAs. ACOAs are said to be disproportionately represented in the human services field. They went to 62 human development conferences in 38 states across the U.S. and handed out 50 questionnaires to anyone who was around after the keynote address. Ackerman and Gondolf got back 1,630 usable questionnaires, or about half of the 3,275 questionnaires they handed out.

One of the questions they asked was: "When you were growing up, did you have a parent who drank too much?" Ackerman and Gondolf randomly selected 500 respondents who answered yes and 500 who answered no to construct their sample of ACOAs and non-ACOAs for data analysis. The ACOAs were much more likely to have witnessed child abuse and spouse abuse than were non-ACOAs, and as adults, the ACOAs had a much higher divorce rate than did non-ACOAs. Ackerman and Gondolf concluded that the data from their purposive sample support the suggestion of clinicians that people who have an alcoholic parent are more likely to have other indicators of a tough family life.

If you think Ackerman and Gondolf had a big purposive sample, consider the work of Kail et al. (1995). They studied female drug addicts who traded sex for money or drugs and female addicts who didn't. The team of researchers had a purposive sample of over 9,000 female drug addicts who were not in treatment. Prostitutes were more likely than nonprostitutes to share needles with others and were less likely to use new needles or to clean old needles before each episode. The sample of female drug addicts was not unbiased; it was, in fact, *intentionally* biased to get answers to questions of practical importance.

Convenience or Haphazard Sampling

Convenience sampling is nothing more than grabbing whoever will stand still long enough to answer your questions. It is useful for exploratory research, to get a feel for "what's going on out there," and for pretesting questionnaires to make sure that the items are unambiguous and not too threatening. In other situations, convenience sampling can be just plain dangerous.

If you ask students at the library how they feel about some current campus issue, you may get different answers than if you ask students who are playing cards in the cafeteria. If you only do interviews around noon, when it is convenient for you, you'll miss all those people for whom noon is not a convenient hour. If you want to know the effect of a new road on some peasants and you only interview people who come to town on the road, you'll miss all the people who live too far off the road for it to do them any good.

It is not necessary to list all the ways that your own prejudices can inflict mortal damage on a convenience sample. Just remember that all samples are representative of *something*. The trick is to make them representative of what *you* want them to be. That's the difference between purposive and convenience sampling.

Still, convenience samples are useful in many situations. Pilot studies are often done with convenience samples. S. Miller et al. (1997) wanted to get a handle on sources of conflict between certified nurse-midwives and physicians who are in collaborative practices. Miller et al. posted a copy of their survey on an electronic bulletin board maintained by a school of nursing and got a convenience sample of 78 nurse-midwife respondents.

The data from these respondents produced a list of common sources of conflict in collaborative practices between nurse-midwives and physicians. We wouldn't put much stock in the fact that a specific *percentage* of the nurse-midwives report conflict with their physician partners over billing of insurance companies for services, but the *list* of conflicts is very instructive because it is the basis for more in-depth research.

Sometimes, convenience samples are all that's available, and you just have to make do. Studies of the homeless in America, for example, are usually done with convenience samples, for obvious reasons. (Burnam and Koegel [1988], however, showed that probability samples are better than convenience samples for estimating how many people actually sleep on the street.) Matthey et al. (1997) compared various tests for identifying major depression among Arabic-speaking and Vietnamese-speaking women in Australia. They interviewed a convenience sample of a few women from each group about the cultural appropriateness of each of the tests. Shih (1997) studied a convenience sample of 35 Taiwanese and Chinese American patients who went through cardiac surgery. Shih interviewed the patients in the intensive care unit during their recovery, clearly a case where a convenience sample is called for.

Snowball Sampling

In snowball sampling, you locate one or more key individuals and ask them to name others who would be likely candidates for your research. Snowball sampling is used in studies of social networks, where the object is to find out who people know and how they know each other. It is used in studies of difficult-to-find populations. E. Miller (1986), for example, used the method to locate female petty criminals and prostitutes in her study of street women.

But "difficult to find" doesn't just mean "people who'd rather not be exposed." It also means any small population for which it is impossible to construct a sampling frame. Richardson (1988), for example, used snowball sampling to locate single or divorced women who were in long-term relationships (a year or more) with married men. A national survey in Britain of 1,997 adults failed to turn up a single self-identified vegan; Beardsworth and Keil (1992) used snowball sampling to find and interview 76 vegetarians in Britain, including 18 strict vegans.

Kadushin (1968) laid out the snowball method for the study of elites. Using key informants and documents, you first construct a preliminary list of elites. These might be "people in this town whose opinions really count" or "living artists whose work everyone wants to buy" or "fellow physicians whose opinions you trust when it comes to adopting a new drug." The elite group can range from very local (a single high school) to international (opinion makers).

Once you have a preliminary list, you show it to several people who are on the list and ask them to name others who they think should be on the list. The process continues until the list becomes "saturated"—that is, until no new names are offered. Elites are easy to find, but hard to interview. Snowball interviewing can be a way out of that problem. Doors open when one member of an elite group passes you on to another.

Ostrander (1980) used snowball sampling to locate informants in her study of class consciousness among upper-class women in a midwestern U.S. city. She selected her first informant by looking for someone who had graduated from an elite women's college, was listed in the social register,

was active in upper-class clubs—and who would talk to her. At the end of the interview, she asked the informant to "suggest another woman of your social group, with a background like yours, who might be willing to talk to me."

David Griffith and his colleagues used two snowball samples in their study of food preferences in Moberly, Missouri. They chose an initial "seed" household in a middle-income neighborhood and asked a man in the house to name three people in town with whom he interacted on a regular basis. The first person cited by the informant lived in a lower-income neighborhood across town. That person, in turn, named other people who were in the lower-income bracket.

After a while, the researchers realized that, though they'd started with a middle-income informant who had children at home, they were getting mostly lower-income, elderly people in the snowball sample. So they started again, this time with a seed from an elite, upper-middle-income neighborhood. By the time they got through, Griffith et al. had a well-balanced sample of 30 informants with whom they did in-depth interviews (reported in Johnson 1990:78).

If you are dealing with a relatively small population of people who are likely to be in contact with one another, then snowball sampling is an effective way to build an exhaustive sampling frame. But in a large population, people who are better known have a better chance of being named in a snowball procedure than people who are less well known. In large populations, then, every person does not have the same chance of being included in a snowball sample.

AND FINALLY

Particularly in ethnographic research, you learn in the field, as you go along, to select the units of analysis (people, court records, etc.) that will provide the information you need. This is what Russell Belk and his colleagues (1988) did in their detailed ethnographic study of buyers and sellers at a swap meet. When you study a process, like bargaining over goods, and you're doing the research in the field, in real time (not under simulated conditions in a lab), then selecting informants who meet certain criteria is the right thing to do.

Good measurement is the key to internal validity, and representative sampling is the key to external validity.

The credibility of research results comes from the power of the methods used in measurement and sampling. Good measurement is the key to internal validity, and representative sampling is the key to external validity. Well-done nonprobability sampling is actually part of good measurement. It contributes to credibility by contributing to internal validity. When someone reads a research report based on really good measurement of a nonprobability sample, they come away thinking, "Yep, I believe those conclusions about the people who were studied in that piece of research."

That's plenty. If you want the credibility of your conclusions to extend beyond the group of people (or countries, organizations, or comic books) you studied, then either (1) repeat the study one or more times with nonprobability samples, or (2) use a probability sample.

Chapter 5 Review

KEY CONCEPTS

Individual data vs. cultural data, 144
Population parameters, 144
Decennial census, 145
Random sample, 146
Panel study, 146
Probability sampling, 147
Nonprobability sampling, 147, 174
Sampling frame, 147
Population, 147
Unbiased sample, 147
Simple random sample, 148
Systematic random sample, 149
Sampling interval, 149
Periodicity, 150
Stratified random sample, 151
Maximizing the between-group variance, 151, 158-160
Minimizing the within-group variance, 151
Sampling design, 151
Age cohorts, 151
Proportionate stratified random sample, 152
Disproportionate stratified random sample, 152
Weighting results, 153-154
Cluster sample, 154-155
Probability proportionate to size, or PPS, 155
Street- and mall-intercept surveys, 157
Sample accuracy, 160
Sample precision, 160
Sample statistics, 161
Parameter estimators, 161
Heterogeneity, 161
Factorial design, 161
Probability distribution, 162
Uniform probability distribution, 162
Exponential distribution, 162
Bimodal distribution, 162
Skewed distribution, 162
Leptokurtic/platykurtic, 162
Normal distribution, 163
Standard deviation, 163
z-scores, or standard scores, 163
Central limit theorem, 164-166
Cumulative mean, 165
Sampling distribution, 165
Standard deviation, 166
Standard error, 167
Confidence interval, 168
Alpha level, 168
Critical region, 168
t-scores, 170
Critical value, 171
Finite population correction, 173
Quota sampling, 175-176
Purposive/judgment sampling, 176-178
Pilot studies, 176
Convenience/haphazard sampling, 178-179
Snowball sampling, 179-180

Chapter 5 Summary

SUMMARY

- There are two kinds of data of interest to social scientists: individual data and cultural data. These two kinds of data require different approaches to sampling.
  - Individual data are about attributes of individuals in a population. To estimate the parameters of these attributes in a population requires probability sampling.
  - Cultural data require experts, which means relying on nonprobability sampling.
- There are several ways to take probability samples.
  - Simple random sampling involves generating a list of random numbers and applying that list to a numbered sampling frame. (Most researchers actually take systematic, rather than simple, random samples.)
  - Stratified random samples are used to ensure that key subpopulations are included in a study. Disproportionate stratified random sampling is used to ensure that important, but relatively small, subpopulations are included in a sample.
  - Cluster sampling is used when there is no overall sampling frame. Cluster sampling is based on the fact that people live in natural clusters (counties, states, etc.) and they participate in the activities of institutions (schools, churches, credit unions, etc.).
  - The best estimates of a parameter are produced in samples taken from clusters of equal size. When clusters are not equal in size, then samples should be taken PPS—with probability proportionate to size.
- Sample size depends on (1) the heterogeneity of the population from which you choose the elements, (2) how many population subgroups you want to deal with simultaneously in your analysis, (3) the size of the phenomenon that you're trying to detect, and (4) how precise you want your parameter estimators to be. Precision involves sampling theory.
  - Sampling theory is based on the normal distribution and the central limit theorem. According to the central limit theorem, if you take many samples of a population, and if the samples are big enough, then (1) the mean and the standard deviation of the sample means will approximate the true mean and standard deviation of the population, and (2) the distribution of sample means will approximate a normal distribution.
  - Sample precision refers to the size of the standard error you are willing to tolerate. As the standard error goes down, the confidence limits become more precise.
- It is often impossible to do strict probability sampling under real research conditions. In these cases, use a nonprobability sample. Also, when you are collecting cultural data, rather than individual-attribute data, random sampling is inappropriate. Some types of nonprobability sampling are quota sampling, purposive/judgment sampling, convenience/haphazard sampling, and snowball sampling.

- In quota sampling, you decide on the subpopulations of interest and on the proportions of those subpopulations in the final sample. Quota sampling resembles stratified probability sampling, but respondents are not chosen randomly. Many commercial polling companies use quota samples that are fine-tuned on the basis of decades of research.
- In purposive, or judgment, sampling you decide the purpose you want the units of analysis (people, communities, countries) to serve. This is somewhat like quota sampling, except that there is no overall sampling design that tells you how many of each type of informant you need for a study.
- Convenience, or haphazard, sampling means grabbing whoever will stand still long enough to answer your questions. It is useful for exploratory research, to get a feel for "what's going on out there," and for pretesting questionnaires to make sure that the items are unambiguous and not too threatening. Pilot studies are often done with convenience samples.
- In snowball sampling, you locate one or more key individuals and ask them to name others who would be likely candidates for your research. Snowball sampling is used in studies of social networks, where the object is to find out who people know and how they know each other. It is used in studies of difficult-to-find populations.