Babbie, Earl. (1995). The Practice of Social Research. 7th Edition. Belmont: Wadsworth, pp. 206-226.

Review of Populations and Sampling Frames

Surprisingly little attention has been given to the issues of populations and sampling frames in social research literature. With this in mind, I've devoted special attention to them here. To further emphasize the point, here is a summary of the main guidelines to remember:

1. Findings based on a sample can be taken as representative only of the aggregation of elements that compose the sampling frame.

2. Often, sampling frames do not truly include all the elements that their names might imply. Omissions are almost inevitable. Thus a first concern of the researcher must be to assess the extent of the omissions and to correct them if possible. (Realize, of course, that the researcher may feel he or she can safely ignore a small number of omissions that cannot easily be corrected.)

3. Even to generalize to the population composing the sampling frame, it is necessary for all elements to have equal representation in the frame: Typically, each element should appear only once. Elements that appear more than once will have a greater probability of selection, and the sample will, overall, overrepresent those elements.

Other, more practical matters relating to populations and sampling frames will be treated elsewhere in this book. For example, the form of the sampling frame—such as a list in a publication, a 3 × 5 card file, mailing address plates, machine-readable cards, or magnetic tapes—can affect how easy it is to use. And ease of use may often take priority over scientific considerations: An "easier" list may be chosen over a "harder" one, even though the latter is more appropriate to the target population. We should not take a dogmatic position in this regard, but every researcher should carefully weigh the relative advantages and disadvantages of such alternatives.

Types of Sampling Designs

Up to this point, we have focused on simple random sampling (SRS). And, indeed, the body of statistics typically used by social researchers assumes such a sample. As we shall see shortly, however, you have a number of options in choosing your sampling method, and you will seldom if ever choose simple random sampling. There are two reasons for that. First, with all but the simplest sampling frame, simple random sampling is not feasible. Second, and probably surprisingly, simple random sampling may not be the most accurate method available. Let's turn now to a discussion of simple random sampling and the other options available.

Simple Random Sampling

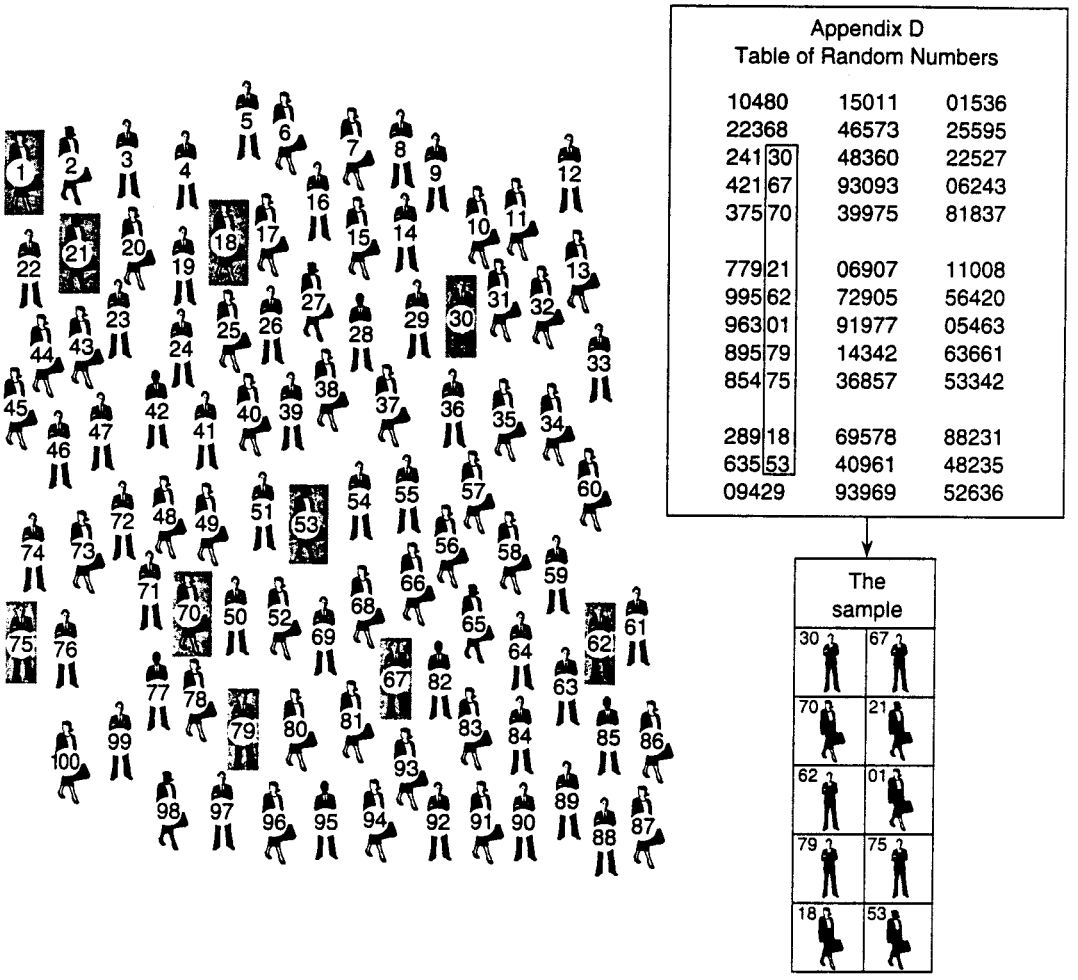
As noted, *simple random sampling* is the basic sampling method assumed in the statistical computations of social research. The mathematics of random sampling are especially complex, so we'll detour around them in favor of describing the ways of employing this method in the field.

Once a sampling frame has been established in accord with the preceding discussion, to use *simple random sampling* the researcher assigns a single number to each element in the list, not skipping any number in the process. A table of random numbers (Appendix D) is then used to select elements for the sample. The box entitled "Using a Table of Random Numbers" explains its use.

If your sampling frame is in a machine-readable form—e.g., computer disk or magnetic tape—a simple random sample can be selected automatically by computer. (In effect, the computer program numbers the elements in the sampling frame, generates its own series of random numbers, and prints out the list of elements selected.)

Figure 8-10 offers a graphic illustration of simple random sampling. Note that the members of our hypothetical micropopulation have

Figure 8-10 A Simple Random Sample



been numbered from 1 to 100. Moving to Appendix D, we decide to use the last two digits of the first column and to begin with the third number from the top. This yields person number 30 as the first one selected into the sample. Number 67 is next, and so forth. (Person 100 would have been selected if "00" had come up in the list.)

Systematic Sampling

Simple random sampling is seldom used in practice. As we shall see, it is not usually the most efficient method, and it can be laborious if done manually. SRS typically requires a list of elements. When such a list is available, researchers usually employ systematic sampling rather than simple random sampling.

In *systematic sampling*, every *k*th element in the total list is chosen (systematically) for inclusion in the sample. If the list contains 10,000 elements and you want a sample of 1,000, you select every tenth element for your sample. To ensure against any possible human bias in using this method, you should select the first element at random. Thus, in the preceding example, you would begin by selecting a random number

Using a Table of Random Numbers

In social research, it is often appropriate to select a set of random numbers from a table such as the one presented in Appendix D. Here's how to do that.

Suppose you want to select a simple random sample of 100 people (or other units) out of a population totaling 980.

1. To begin, number the members of the population: in this case, from 1 to 980. Now the problem is to select 100 random numbers. Once you've done that, your sample will consist of the people having the numbers you've selected. (*Note:* It's not essential to actually number them, as long as you're sure of the total. If you have them in a list, for example, you can always count through the list after you've selected the numbers.)

2. The next step is to determine the number of digits you will need in the random numbers you select. In our example, there are 980 members of the population, so you will need three-digit numbers to give everyone a chance of selection. (If there were 11,825 members of the population, you'd need to select five-digit numbers.) Thus, we want to select 100 random numbers in the range from 001 to 980.

3. Now turn to the first page of Appendix D. Notice there are several rows and columns of five-digit numbers, and there are several pages. The table represents a series of random numbers in the range from 00001 to 99999. To use the table for your hypothetical sample, you have to answer these questions:

   a. How will you create three-digit numbers out of five-digit numbers?
   b. What pattern will you follow in moving through the table to select your numbers?
   c. Where will you start?

Each of these questions has several satisfactory answers. The key is to create a plan and follow it. Here's an example.

4. To create three-digit numbers from five-digit numbers, let's agree to select five-digit numbers from the table but consider only the left-most three digits in each case. If we picked the first number on the first page—10480—we would only consider the 104. (We could agree to take the digits furthest to the right, 480, or the middle three digits, 048, and any of these plans would work.) The key is to make a plan and

between one and ten. The element having that number is included in the sample, plus every tenth element following it. This method is technically referred to as a *systematic sample with a random start*. Two terms are frequently used in connection with systematic sampling. The **sampling interval** is the standard distance between elements selected in the sample: ten in the preceding sample. The **sampling ratio** is the proportion of elements in the population that are selected: 1/10 in the example.

$$\text{sampling interval} = \frac{\text{population size}}{\text{sample size}}$$

$$\text{sampling ratio} = \frac{\text{sample size}}{\text{population size}}$$

In practice, systematic sampling is virtually identical to simple random sampling. If the list of elements is indeed randomized before sampling, one might argue that a systematic sample drawn from that list is in fact a simple random sample. By now, debates over the relative merits of simple random sampling and systematic sampling have been resolved largely in favor of the simpler method: systematic sampling. Empirically, the results are virtually identical. And, as we shall see in a later section, systematic

stick with it. For convenience, let's use the left-most three digits.

5. We can also choose to progress through the tables any way we want: down the columns, up them, across to the right or to the left, or diagonally. Again, any of these plans will work just fine so long as we stick to it. For convenience, let's agree to move down the columns. When we get to the bottom of one column, we'll go to the top of the next; when we exhaust a given page, we'll start at the top of the first column of the next page.

6. Now, where do we start? You can close your eyes and stick a pencil into the table and start wherever the pencil point lands. (I know it doesn't sound scientific, but it works.) Or, if you're afraid you'll hurt the book or miss it altogether, close your eyes and make up a column number and a row number. ("I'll pick the number in the fifth row of column 2.") Start with that number.

7. Let's suppose we decide to start with the fifth number in column 2. If you look on the first page of Appendix D, you'll see that the starting number is 39975. We have selected 399 as our first random number, and we have 99 more to go. Moving down the second column, we select 069, 729, 919, 143, 368, 695, 409, 939, and so forth. At the bottom of column 2, we select number 104 and continue to the top of column 3: 015, 255, and so on.

8. See how easy it is? But trouble lies ahead. When we reach column 5, we are speeding along, selecting 816, 309, 763, 078, 061, 277, 988 . . . Wait a minute! There are only 980 students in the senior class. How can we pick number 988? The solution is simple: Ignore it. Any time you come across a number that lies outside your range, skip it and continue on your way: 188, 174, and so forth. The same solution applies if the same number comes up more than once. If you select 399 again, for example, just ignore it the second time.

9. That's it. You keep up the procedure until you've selected 100 random numbers. Returning to your list, your sample consists of person number 399, person number 69, person number 729, and so forth.

sampling, in some instances, is slightly more accurate than simple random sampling.

There is one danger involved in systematic sampling. The arrangement of elements in the list can make systematic sampling unwise. Such an arrangement is usually called *periodicity*. If the list of elements is arranged in a cyclical pattern that coincides with the sampling interval, a grossly biased sample may be drawn. Two examples will illustrate.

In one study of soldiers during World War II, the researchers selected a systematic sample from unit rosters. Every tenth soldier on the roster was selected for the study. The rosters, however, were arranged in a table of organizations: sergeants first, then corporals and privates, squad by squad. Each squad had ten members. As a result, every tenth person on the roster was a squad sergeant. The systematic sample selected contained only sergeants. It could, of course, have been the case that no sergeants were selected for the same reason.

As another example, suppose we select a sample of apartments in an apartment building. If the sample is drawn from a list of apartments arranged in numerical order (for example, 101, 102, 103, 104, 201, 202, and so on), there is a danger of the sampling interval coinciding with

the number of apartments on a floor or some multiple thereof. Then the samples might include only northwest-corner apartments or only apartments near the elevator. If these types of apartments have some other particular characteristic in common (for example, higher rent), the sample will be biased. The same danger would appear in a systematic sample of houses in a subdivision arranged with the same number of houses on a block.

In considering a systematic sample from a list, then, you should carefully examine the nature of that list. If the elements are arranged in any particular order, you should figure out whether that order will bias the sample to be selected and take steps to counteract any possible bias (for example, take a simple random sample from cyclical portions).

In summary, however, systematic sampling is usually superior to simple random sampling, in convenience if nothing else. Problems in the ordering of elements in the sampling frame can usually be remedied quite easily.

Stratified Sampling

In the two preceding sections, we discussed two methods of sample selection from a list: random and systematic. **Stratification** is not an alternative to these methods, but it represents a possible modification in their use.

Simple random sampling and systematic sampling both ensure a degree of representativeness and permit an estimate of the error present. Stratified sampling is a method for obtaining a greater degree of representativeness—decreasing the probable sampling error. To understand why that is the case, we must return briefly to the basic theory of sampling distribution.

We recall that sampling error is reduced by two factors in the sample design. First, a large sample produces a smaller sampling error than a small sample. Second, a homogeneous population produces samples with smaller sampling errors than does a heterogeneous population. If 99 percent of the population agrees with a certain statement, it is extremely unlikely that any probability sample will greatly misrepresent the extent of agreement. If the population is split 50–50 on the statement, then the sampling error will be much greater.

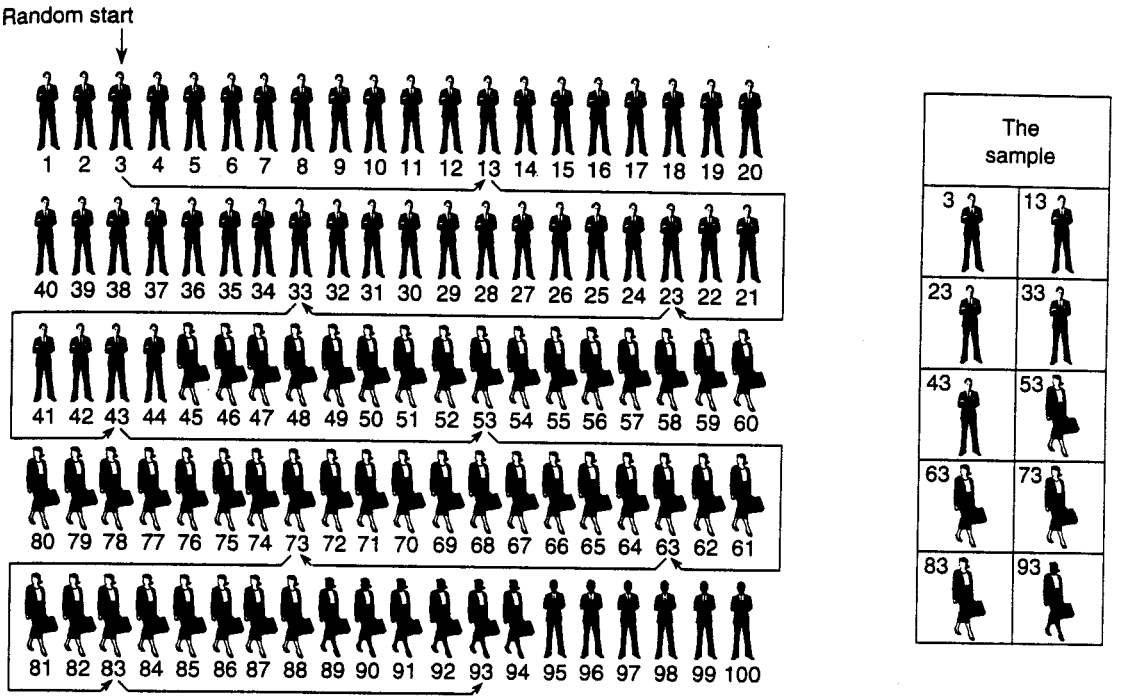
Stratified sampling is based on this second factor in sampling theory. Rather than selecting your sample from the total population at large, you ensure that appropriate numbers of elements are drawn from homogeneous subsets of that population. To get a stratified sample of university students, for example, you would first organize your population by college class and then draw appropriate numbers of freshmen, sophomores, juniors, and seniors. In a nonstratified sample, representation by class would be subjected to the same sampling error as other variables. In a sample stratified by class, the sampling error on this variable is reduced to zero.

Even more complex stratification methods are possible. In addition to stratifying by class, you might also stratify by sex, by grade point average, and so forth. In this fashion you might be able to ensure that your sample would contain the proper numbers of freshman males with a 4.0 average, of freshman females with a 4.0 average, and so forth.

The ultimate function of stratification, then, is to organize the population into homogeneous subsets (with heterogeneity between subsets) and to select the appropriate number of elements from each. To the extent that the subsets are homogeneous on the stratification variables, they may be homogeneous on other variables as well. Because age is related to college class, a sample stratified by class will be more representative in terms of age as well. Because occupational aspirations still seem to be related to sex, a sample stratified by sex will be more representative in terms of occupational aspirations.

The choice of stratification variables typically depends on what variables are available. Sex can often be determined in a list of names. University lists are typically arranged by class. Lists of faculty members may indicate their departmental affiliation. Government agency files

Figure 8-11 A Stratified, Systematic Sample with a Random Start



may be arranged by geographical region. Voter registration lists are arranged according to precinct.

In selecting stratification variables from among those available, however, you should be concerned primarily with those that are presumably related to variables that you want to represent accurately. Because sex is related to many variables and is often available for stratification, it is often used. Education is related to many variables, but it is often not available for stratification. Geographical location within a city, state, or nation is related to many things. Within a city, stratification by geographical location usually increases representativeness in social class, ethnic group, and so forth. Within a nation, it increases representativeness in a broad range of attitudes as well as in social class and ethnicity.

Methods of stratification in sampling vary. When you are working with a simple list of all elements in the population, two are predominant. One method is to sort the population elements into discrete groups based on whatever stratification variables are being used. On the basis of the relative proportion of the population represented by a given group, you select—randomly or systematically—a number of elements from that group constituting the same proportion of your desired sample size. For example, if freshman men with a 4.0 average compose 1 percent of the student population and you desire a sample of 1,000 students, you would select 10 freshman men with a 4.0 average.

The other method is to group students as described and then put those groups together in a continuous list, beginning with all freshman men with a 4.0 average and ending with all senior women with a 1.0 or below. You would then select a systematic sample, with a random start, from the entire list. Given the arrangement of the list, a systematic sample would

select proper numbers (within an error range of 1 or 2) from each subgroup. (*Note:* A simple random sample drawn from such a composite list would cancel out the stratification.)

Figure 8-11 offers a graphic illustration of stratified, systematic sampling. As you can see, we lined up our micropopulation according to sex and race. Then, beginning with a random start of "3," we've taken every tenth person thereafter: 3, 13, 23, . . . , 93.

Stratified sampling ensures the proper representation of the stratification variables to enhance representation of other variables related to them. Taken as a whole, then, a stratified sample is likely to be more representative on a number of variables than a simple random sample. Although the simple random sample is still regarded as somewhat sacred, it should now be clear that you can often do better.

Implicit Stratification in Systematic Sampling

I mentioned that systematic sampling can, under certain conditions, be more accurate than simple random sampling. That is the case whenever the arrangement of the list creates an implicit stratification. As already noted, if a list of university students is arranged by class, then a systematic sample provides a stratification by class where a simple random sample would not.

In a study of students at the University of Hawaii, after stratification by school class, the students were arranged by their student identification numbers. These numbers, however, were their social security numbers. The first three digits of the social security number indicate the state in which the number was issued. As a result, within a class, students were arranged by the state in which they were issued a social security number, providing a rough stratification by geographical origins.

You should realize, therefore, that an ordered list of elements may be more useful to you than an unordered, randomized list. This point has been stressed here in view of an unfortunate belief that lists should be randomized before systematic sampling. Only if the arrangement presents the problems discussed earlier should the list be rearranged.

Illustration: Sampling University Students

Let's put these principles into practice by looking at an actual sampling design used to select a sample of university students. The purpose of the study was to survey, with a mail-out questionnaire, a representative cross section of students attending the main campus of the University of Hawaii in 1968. The following sections will describe the steps and decisions involved in selecting that sample.

Study Population and Sampling Frame

The obvious sampling frame available for use in this sample selection was the computerized file maintained by the university administration. The tape contained students' names, local and permanent addresses, social security numbers, and a variety of other information such as field of study, class, age, and sex.

The computer database, however, contained files on all persons who could, by any conceivable definition, be called students, many of whom seemed inappropriate to the purposes of the study. As a result, it was necessary to define the *study population* in a somewhat more restricted fashion. The final definition included those 15,225 day program degree candidates registered for the fall 1968 semester on the Manoa campus of the university, including all colleges and departments, both undergraduate and graduate students, and both American and foreign students. The computer program used for sampling, therefore, limited consideration to students fitting this definition.

Stratification

The sampling program also permitted stratification of students before sample selection. The researchers decided that stratification by college class would be sufficient, although the students might have been further stratified within class, if desired, by sex, college, major, and so forth.

Sample Selection

Once the students had been arranged by class, a systematic sample was selected across the entire rearranged list. The sample size for the study was initially set at 1,100. To achieve this sample, the sampling program was set for a 1/14 sampling ratio. The program generated a random number between 1 and 14; the student having that number and every fourteenth student thereafter was selected in the sample.

Once the sample had been selected, the computer was instructed to print each student's name and mailing address on six self-adhesive mailing labels. These labels were then simply transferred to envelopes for mailing the questionnaires.

Sample Modification

This initial design of the sample had to be modified. Before the mailing of questionnaires, the researchers discovered that unexpected expenses in the production of the questionnaires made it impossible to cover the costs of mailing to all 1,100 students. As a result, one-third of the mailing labels were systematically selected (with a random start) for exclusion from the sample. The final sample for the study was thereby reduced to about 770.

This modification to the sample is mentioned here to illustrate the frequent necessity to change aspects of the study plan in midstream. Because the excluded students were systematically omitted from the initial systematic sample, the remaining 770 students could still be taken as reasonably representing the study population. The reduction in sample size did, of course, increase the range of sampling error.

Multistage Cluster Sampling

The preceding sections have dealt with reasonably simple procedures for sampling from lists of elements. Such a situation is ideal. Unfortunately, however, much interesting social research requires the selection of samples from populations that cannot be easily listed for sampling purposes. Examples would be the population of a city, state, or nation; all university students in the United States; and so forth. In such cases, the sample design must be much more complex. Such a design typically involves the initial sampling of *groups* of elements—clusters—followed by the selection of elements within each of the selected clusters.

Cluster sampling may be used when it is either impossible or impractical to compile an exhaustive list of the elements composing the target population. All church members in the United States would be an example of such a population. It is often the case, however, that the population elements are already grouped into subpopulations, and a list of those subpopulations either exists or can be created practically. Thus, church members in the United States belong to discrete churches, and it would be possible to discover or create a list of those churches. Following a cluster sample format, then, the list of churches would be sampled in some manner as discussed previously (for example, a stratified, systematic sample). Next, you would obtain lists of members from each of the selected churches. Each of the lists would then be sampled, to provide samples of church members for study. (For an example, see Glock, Ringer, and Babbie, 1967.)

Another typical situation concerns sampling among population areas such as a city. Although there is no single list of a city's population,

citizens reside on discrete city blocks or census blocks. It is possible, therefore, to select a sample of blocks initially, create a list of persons living on each of the selected blocks, and subsample persons on each block.

In a more complex design, you might sample blocks, list the households on each selected block, sample the households, list the persons residing in each household, and, finally, sample persons within each selected household. This multistage sample design would lead to the ultimate selection of a sample of individuals but would not require the initial listing of all individuals in the city's population.

Multistage cluster sampling, then, involves the repetition of two basic steps: listing and sampling. The list of primary sampling units (churches, blocks) is compiled and, perhaps, stratified for sampling. Then a sample of those units is selected. The selected primary sampling units are then listed and perhaps stratified. The list of secondary sampling units is then sampled, and so forth.

Multistage cluster sampling makes possible studies that would otherwise be impossible. Consider, for example, the "Santa Claus" survey described in the box entitled "Sampling Santa's Fans."

Cluster sampling is highly recommended by its efficiency, but the price of that efficiency is a less accurate sample. A simple random sample drawn from a population list is subject to a single sampling error, but a two-stage cluster sample is subject to two sampling errors. First, the initial sample of clusters will represent the population of clusters only within a range of sampling error. Second, the sample of elements selected within a given cluster will represent all the elements in that cluster only within a range of sampling error. Thus, for example, you run a certain risk of selecting a sample of disproportionately wealthy city blocks, plus a sample of disproportionately wealthy households within those blocks. The best solution to this problem lies in the number of clusters selected initially and the number of elements selected within each.

Typically, you'll be restricted to a total sample size; for example, you may be limited to conducting 2,000 interviews in a city. Given this broad limitation, however, you have several options in designing your cluster sample. At the extremes you might choose one cluster and select 2,000 elements within that cluster; or you might select 2,000 clusters with one element selected within each. Of course, neither of these extremes is advisable, but a broad range of choices lies between them. Fortunately, the logic of sampling distributions provides a general guideline to be followed.

Recall that sampling error is reduced by two factors: an increase in the sample size and increased homogeneity of the elements being sampled. These factors operate at each level of a multistage sample design. A sample of clusters will best represent all clusters if a large number are selected and if all clusters are very much alike. A sample of elements will best represent all elements in a given cluster if a large number are selected from the cluster and if all the elements in the cluster are very much alike.

With a given total sample size, however, if the number of clusters is increased, the number of elements within a cluster must be decreased. In this respect, the representativeness of the clusters is increased at the expense of more poorly representing the elements composing each cluster, or vice versa. Fortunately, the factor of homogeneity can be used to ease this dilemma.

Typically, the elements composing a given natural cluster within a population are more homogeneous than are all elements composing the total population. The members of a given church are more alike than are all church members; the residents of a given city block are more alike than are all the residents of a whole city. As a result, relatively fewer elements may be needed to adequately represent a given natural cluster, although a larger number of clusters may be needed to adequately represent the diversity found among the clusters. This fact is most clearly seen in the extreme case of very

Sampling Santa's Fans

With the approach of Christmas, 1985, *The New York Times* thought it would be interesting to survey the nation's children regarding their beliefs in Santa Claus. There being no national registry of those who've been naughty and nice, the *Times* had to use some ingenuity. Here's their description of what they did:

The latest New York Times Poll is based on telephone interviews conducted Dec. 14–18 with 261 children ages 3 through 10 around the United States, excluding Alaska and Hawaii.

The sample of telephone exchanges called was selected by a computer from a complete list of exchanges in the country. The exchanges were chosen to ensure that each region of the country was represented in proportion to its population. For each exchange, the telephone numbers were formed by random digits, thus permitting access to both listed and unlisted residential numbers.

After interviews with 1,358 adults were completed, parents were asked if their children could be interviewed on the subject of Christmas. The results have been weighted to take account of household size and number of residential telephones and to adjust for variations in the sample relation to region, race, sex, age and education.

By the way, 87 percent of the children said they believed in Santa Claus: ranging from 96 percent among those 3–5 down to 69 percent among the 9–10 year-olds.

Source: Sara Rimer, "Poll Sees Landslide for Santa: Of U.S. Children, 87% Believe," *The New York Times*, December 24, 1985.

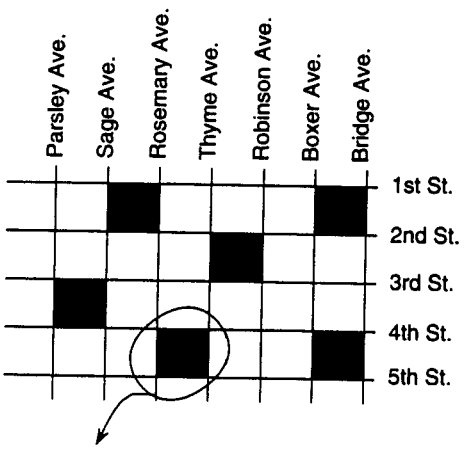
different clusters composed of identical elements within each. In such a situation, a large number of clusters would adequately represent all its members. Although this extreme sitution never exists in reality, it is closer to the truth in most cases than its opposite: identical clusters composed of grossly divergent elements.

The general guideline for cluster design, then, is to maximize the number of clusters selected while decreasing the number of elements within each cluster. It must be noted, however, that this scientific guideline must be balanced against an administrative constraint. The efficiency of cluster sampling is based on the ability to minimize the listing of population elements. By initially selecting clusters, you need only list the elements composing the selected clusters, not all elements in the entire population. Increasing the number of clusters, however, goes directly against this efficiency factor in cluster sampling. A small number of clusters may be listed more quickly and more cheaply than a large number. (Remember that all the elements in a selected cluster must be listed even if only a few are to be chosen in the sample.)

The final sample design will reflect these two constraints. In effect, you will probably select as many clusters as you can afford. Lest this issue be left too open-ended at this point, one rule of thumb may be presented. Population researchers conventionally aim for the selection of 5 households per census block. If a total of 2,000 households are to be interviewed, you would aim at 400 blocks with 5 household interviews on each. Figure 8-12 presents a graphic overview of this process.

Before turning to more detailed procedures available to cluster sampling, it bears repeating that this method almost inevitably involves a loss of accuracy. The manner in which this appears, however, is somewhat complex. First, as noted earlier, a multistage sample design is subject to a sampling error at each stage. Because the sample size is necessarily smaller at each stage than the total sample size, the sampling error at each stage will be greater than would

Figure 8-12 Multistage Cluster Sampling



Stage One: Identity blocks and select a sample. (Selected blocks are shaded.)

Stage Two: Go to each selected block and list all households in order. (Example of one listed block.)

1. 491 Rosemary Ave.
2. 487 Rosemary Ave.
3. 473 Rosemary Ave.
4. 455 Rosemary Ave.
5. 437 Rosemary Ave. ••
6. 423 Rosemary Ave.
7. 411 Rosemary Ave.
8. 403 Rosemary Ave.
9. 1101 4th St.
10. 1123 4th St.
11. 1137 4th St. ••
12. 1157 4th St.
13. 1169 4th St.
14. 1187 4th St.
15. 402 Thyme Ave.
16. 408 Thyme Ave.
17. 424 Thyme Ave. ••
18. 446 Thyme Ave.
19. 458 Thyme Ave.
20. 480 Thyme Ave.
21. 498 Thyme Ave.
22. 1186 5th St.
23. 1174 5th St. ••
24. 1160 5th St.
25. 1140 5th St.
26. 1122 5th St.
27. 1118 5th St.
28. 1116 5th St.
29. 1104 5th St. ••
30. 1102 5th St.

Stage Three: For each list, select sample of households. (In this example, every sixth household has been selected starting with #5, which was selected at random.)

be the case for a single-stage random sample of elements. Second, sampling error is estimated on the basis of observed variance among the sample elements. When those elements are drawn from among relatively homogeneous clusters, the estimated sampling error will be too optimistic and must be corrected in the light of the cluster sample design.

Multistage Cluster Sampling, Stratification

Thus far, we have looked at cluster sampling as though a simple random sample were selected at each stage of the design. In fact, stratification techniques can be used to refine and improve the sample being selected.

The basic options available are essentially the same as those possible in single-stage sampling from a list. In selecting a national sample of churches, for example, you might initially stratify your list of churches by denomination, geographical region, size, rural or urban location, and perhaps by some measure of social class.

Once the primary sampling units (churches, blocks) have been grouped according to the relevant, available stratification variables, either simple random or systematic sampling techniques can be used to select the sample. You might select a specified number of units from each group or *stratum*, or you might arrange the stratified clusters in a continuous list and systematically sample that list.

To the extent that clusters are combined into homogeneous strata, the sampling error at this stage will be reduced. The primary goal of stratification, as before, is homogeneity.

There is no reason why stratification could not take place at each level of sampling. The elements listed within a selected cluster might be stratified before the next stage of sampling. Typically, however, that is not done. (Recall the assumption of relative homogeneity within clusters.)

Probability Proportionate to Size (PPS) Sampling

In this section, I want to give you an introduction to a more sophisticated form of cluster sampling used in many large-scale survey sampling projects. In the preceding discussion, I talked about selecting a random or systematic sample of clusters and then a random or systematic sample of elements within each cluster selected. Notice that this produces an overall sampling scheme in which every element in the whole population has the same probability of selection.

Let's say we are selecting households within a city. If there are 1,000 city blocks and we initially select a sample of 100, that means that each block has a 100/1,000 or .1 chance of being selected. If we next select 1 household in 10 from those residing on the selected blocks, each household has a .1 chance of selection within its block. To calculate the overall probability of a household being selected, we simply multiply the probabilities at the individual steps in sampling. That is, each household has a 1/10 chance of its block being selected and a 1/10 chance of that specific household being selected if the block is one of those chosen. Each household, in this case, has a $1/10 \times 1/10 = 1/100$ chance of selection overall. Because each household would have the same chance of selection, the sample so selected should be representative of all households in the city.

There are dangers in this procedure, however. In particular, the varying sizes of blocks (measured in numbers of households) present a problem. Let's suppose that half the city's population resides in 10 densely packed blocks filled with high-rise apartment buildings, and suppose that the rest of the population lives in single-family dwellings spread out over the remaining 900 blocks. When we first select our sample of 1/10 of the blocks, it is quite possible that we'll miss all of the 10 densely packed high-rise blocks. No matter what happens in the second stage of sampling, our final sample of

households will be grossly unrepresentative of the city, being composed only of single-family dwellings.

Whenever the clusters sampled are of greatly differing sizes, it is appropriate to use a modified sampling design called *probability proportionate to size*—**PPS.** This design (1) guards against the problem I've just described and (2) still produces a final sample in which each element has the same chance of selection.

As the name suggests, each cluster is given a chance of selection proportionate to its size. Thus, a city block with 200 households has twice the chance of selection as one with only 100 households. The method for doing this is demonstrated in the illustration that follows in the next section. Within each cluster, however, a fixed *number* of elements is selected, say, 5 households per block. Notice how this procedure results in each household having the same probability of selection overall.

Let's say that Household A is located on a block containing 100 households altogether, and Household B is located on a block containing 25 households. Suppose that we plan to select 5 households from whatever blocks are picked. This means that if the block containing Household A is picked, that household has a 5/100 chance of selection in the second stage of sampling. If the block containing Household B is picked, it has a 5/25 or 20/100 chance of being selected in the second stage of sampling. At the second stage of sampling, then, Household B has four times as good a chance of having its block selected in the first stage as did Household A. In the overall sampling design, then, both Household A and Household B have the same chance of being selected.

In a PPS sample design, the overall probability of selection for elements is calculated as follows:

1. The probability of a cluster being selected is equal to its proportional share of all the elements in the population times the number of clusters to be selected.

2. The probability of an element being selected within a cluster is equal to the number to be selected within each cluster divided by the number of elements contained within that particular cluster.

3. The overall probability of an element being selected equals (1) times (2).

Here's an example. Suppose a city is composed of 2,000 blocks and 100,000 households and that we want to select 1,000 households. That means that each household should have a 1,000/100,000 or .01 chance of selection. We decide to accomplish this by picking 200 blocks PPS and selecting 5 households on each of the blocks chosen.

Now consider a block containing 100 households. The block has a probability of selection equal to:

$$\begin{matrix}200\\ \text{blocks}\\ \text{to be}\\ \text{chosen}\end{matrix} \times \frac{100}{\dfrac{\text{(households on the block)}}{\text{(households in the city)}}} = \frac{100}{100{,}000} = .2$$

If that block is selected, each household has a second-stage probability of selection equal to:

$$\frac{5\text{ (to be selected on each block)}}{100\text{ (households on that block)}} = .05$$

Multiplying .2 times .05, we get an overall probability of selection equal to .01, as required.

Now let's consider a block with only 20 households on it. The block's chance of selection is only 200 × 20/100,000 or .04, much less than the earlier example. If this block is selected, however, each household has a chance of 5/20 or .25 of selection in the second stage. Overall, its probability of selection is .04 times .25, or .01: the same as the earlier case and as required by the overall sample design.

If you examine the method for calculating overall probabilities carefully, you'll see why the result is always going to be the same.

Example 1: 200 × 100/100,000 × 5/100 = .01

Example 2: 200 × 20/100,000 × 5/20 = .01

The only thing that differs in the two examples is the number of households on the blocks, and that number appears in both numerator and denominator, thus canceling itself out. No matter what the block size, then, the overall probability of a household being selected will be equal to 200 times 5/100,000, or .01. See how neat and clean it is?

As a practical matter, PPS cluster sampling is never quite this neat in the field. For example, the estimates of the number of households on blocks are seldom totally accurate, so that blocks are given too high or too low probabilities of selection. There are several statistical solutions to this problem, however, although it's probably not necessary for you to know them for purposes of this introduction to sampling. Other problems arise because some households selected cannot be interviewed—either the people are never home or they refuse to be interviewed. Again, the structure of PPS sampling makes adjustments possible, so that the data collected from households that are selected and interviewed can be taken as representative of all households in the city.

I began this section on PPS sampling by pointing to the danger of missing very large blocks altogether. There is another benefit inherent in this more sophisticated design. If you recall the earlier discussion of homogeneity and heterogeneity, you'll remember that sampling is less of a problem if all the elements being selected are pretty much alike, that is, homogeneous. PPS sampling takes advantage of that fact, in the sense that households composing a single city block or similar geographical groupings are likely to be quite similar to one another, as are the families residing in them. Specifically, the similarity of households on a block is greater than the similarity among households in a whole city. This means that it takes relatively few households on a single block to describe all the households on that block. As a rule of thumb, 5 is usually enough in the context of a large multistage cluster sample. Observing more than 5 households on a single block would improve the description of the block slightly, but the description of the city as a whole would be better improved by adding more blocks to the sample than adding more households on fewer blocks. Given that you can only interview, say, 1,000 households altogether, it would be better to interview 5 each on 200 blocks than to interview 20 each on 50 blocks. In addition to guarding against specific dangers, then, PPS sampling is an efficient use of limited resources.

Disproportionate Sampling and Weighting

Ultimately, a probability sample is representative of a population if all elements in the population have an equal chance of selection in that sample. Thus, in each of the preceding discussions we have noted that the various sampling procedures result in an equal chance of selection—even though the ultimate selection probability is the product of several partial probabilities.

More generally, however, a probability sample is one in which each population element has a *known nonzero* probability of selection—even though different elements may have different probabilities. If controlled probability sampling procedures have been used, any such sample may be representative of the population from which it is drawn if each sample element is assigned a weight equal to the inverse of its probability of selection. Thus, where all sample elements have had the same chance of selection, each is given the same weight: 1. (This is called a *self-weighting* sample.)

Disproportionate sampling and weighting come into play in two basic ways. First, you may sample subpopulations disproportionately to ensure sufficient numbers of cases from each for analysis. For example, a given city may have a suburban area containing one-fourth of its total population. Yet you might be especially interested in a detailed analysis of households in that area and may feel that one-fourth of this total sample size would be too few. As a result, you might decide to select the same number of households from the suburban area as from the

remainder of the city. Households in the suburban area, then, are given a disproportionately better chance of selection than those located elsewhere in the city.

As long as you analyze the two area samples separately or comparatively, you need not worry about the differential sampling. If you want to combine the two samples to create a composite picture of the entire city, however, you must take the disproportionate sampling into account. If n is the number of households selected from each area, then the households in the suburban area had a chance of selection equal to n divided by one-fourth of the total city population. Because the total city population and the sample size are the same for both areas, the suburban-area households should be given a weight of $\frac{1}{4}n$, and the remaining households should be given a weight of $\frac{3}{4}n$. This weighting procedure could be simplified by merely giving a weight of 3 to each of the households selected outside the suburban area. (This procedure gives a proportionate representation to each sample element. The population figure would have to be included in the weighting if population estimates were desired.)

Here's an example of the problems that can be created when disproportionate sampling is not accompanied by a weighting scheme. When the *Harvard Business Review* decided to survey its subscribers on the issue of sexual harassment at work, for example, it seemed appropriate to oversample women. Here's how Collins and Blodgett explained the matter:

> We also skewed the sample another way: to ensure a representative response from women, we mailed a questionnaire to virtually every female subscriber, for a male/female ratio of 68% to 32%. This bias resulted in a response to 52% male and 44% female (and 4% who gave no indication of gender)—compared to HBR's U.S. subscriber proportion of 93% male and 7% female.
>
> (1981:78)

You should have noticed a couple of things in this quotation. First, it would be nice to know a little more about what "virtually every female" means. Evidently, they didn't send questionnaires to all female subscribers, but there's no indication of who was omitted and why. Second, they didn't use the term *representative* in its normal social science usage. What they mean, of course, is that they want to get a substantial or "large enough" response from women, and oversampling is a perfectly acceptable way of accomplishing that.

By sampling more women than a straightforward probability sample would have produced, they have gotten enough women (812) to compare with the men (960). Thus, when the authors report, for example, that 32 percent of the women and 66 percent of the men agree that "the amount of sexual harassment at work is greatly exaggerated," we know that the female response is based on a substantial number of cases. That's good. There are problems, however.

To begin, subscriber surveys are always problematic. In this case, the best the researchers can hope to talk about is "what subscribers to *Harvard Business Review* think." In a loose way, it might make sense to think of that population as representing the more sophisticated portion of corporate management. Unfortunately, the overall response rate was 25 percent. Although that is quite good for subscriber surveys, it is a low response rate in terms of generalizing from probability samples.

Beyond that, however, the disproportionate sample design creates a further problem. When the authors state that 73 percent favor company policies against harassment (Collins and Blodgett, 1981:78), that figure is undoubtedly too high, since the sample contains a disproportionately high percentage of women—who are more likely to favor such policies. And, when the researchers report that top managers are more likely to feel that claims of sexual harassment are exaggerated than are middle- and lower-level management (1981:81), that finding is also suspect. As the researchers report, women are disproportionately represented in lower management. That alone might account for the apparent differences in different levels

of management. In short, the failure to take account of the oversampling of women confounds all survey results that don't separate findings by sex.

Degrees of Precision in Weighting

In any complex sample design, you face a number of options with regard to weighting of purposively or inadvertently disproportionate samples. You may compute weights for each element to several decimal places, or you may assign rough weights to account for only the grossest instances of disproportionate sampling. In the previous case of the city in which the suburban area was oversampled, it is unlikely that the population of that area was exactly one-fourth of the city's population: Suppose it actually was .25001, .2600, or .2816 of the total population. In each of these instances, whether you choose to apply the rough overall weighting of one-fourth or decide to be more precise will depend on the precision you desire in your findings. If your research purposes can tolerate errors of a few percentage points, you will probably not waste your time and effort in weighting exactly. In deciding the degree of precision required, moreover, you should take into account the degree of error to be expected from normal sampling distribution, plus all the various types of nonsampling error.

Ultimately, there is no firm guideline for you to follow in determining the precision to be sought in weighting. As in so many other aspects of study design, you have considerable latitude. At the same time, however, you should bear your decision in mind when reporting your findings. You should not employ only a rough weighting procedure and then suggest that your findings are accurate within a minuscule range of error.

Methods for Weighting

Having outlined the scientific concerns for determining the degree of precision desired in weighting, I should note that the choice will often be made on the basis of available methods for weighting. There are two basic methods for weighting.

1. For the rough weighting of samples drawn from subpopulations, weighted tables can be constructed from the unweighted tables for each of the subsamples. In the earlier example, you could create a raw table of distributions for the suburban sample and for the nonsuburban sample separately, triple the number of cases in each cell of the nonsuburban table, add the cases across the two tables, and compute percentages for the composite table.

2. If the data are to be analyzed by computer, a special program may be designed to assign a precise weight to each case in the data file. This way, the computer does all the work of weighting.

It bears repeating that complex weighting methods are usually not necessary in social research. Understanding the logic of weighting, however, will strengthen your grasp of the logic of sampling in general.

Figure 8-13 offers a graphic illustration of the logic of weighting. You'll recall that our micropopulation of 100 people (Figure 8-1) contains only 12 African Americans—6 men and 6 women (which is the approximate percentage of African Americans in America). Representative samples from such a population would provide few African-American subjects for study, which is a problem if we wished to compare African Americans and whites, for example. We might resolve this problem by oversampling African Americans, then weighting the resulting data.

In the illustration, we've selected all the African Americans into our sample, but only one-fourth of the whites. Having collected the data, we've then weighted each white respondent's data by a factor of four. If we did this by actually making copies of the data files, we would end up with 100 data files—which could be analyzed as though we had collected data from all 100 members of the population.

Figure 8-13 Sampling Disproportionately and Weighting

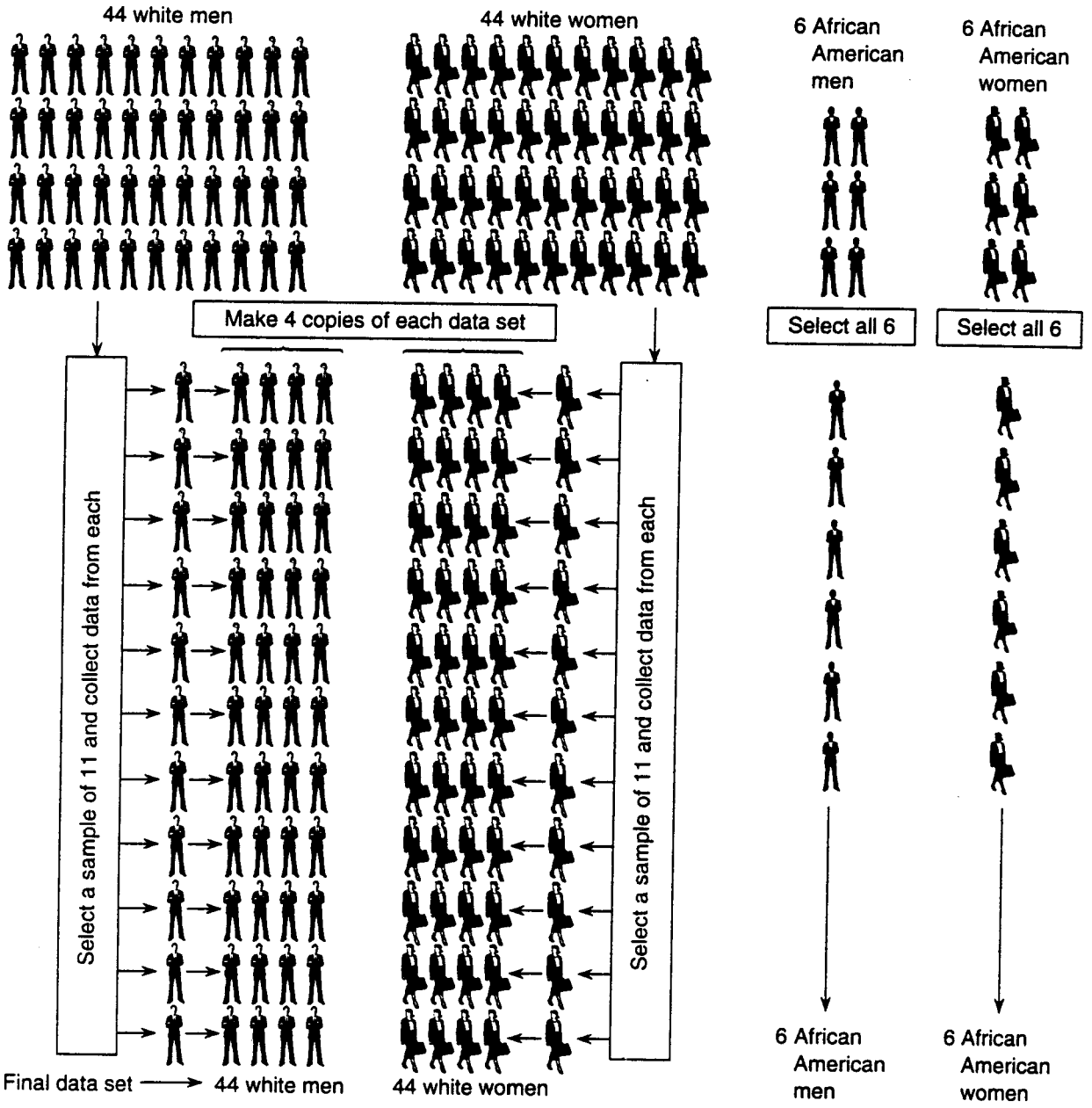


Illustration: Sampling Churchwomen

Let's complete the discussion of cluster sampling with a research example. The illustration that follows is not as complex as the area probability samples that are employed in studies of geographic areas such as cities, states, or the nation. Nonetheless, it should be a useful example of the various principles of cluster sampling.

The purpose of this study was to examine the religious and social attitudes of women members of churches in a diocese of the Episcopal church. A representative sample of all churchwomen in the diocese was desired. Because there was no single list of such women, a multi-

stage sample design was created. In the initial stage of sampling, churches were selected with probability proportionate to size (PPS), and then women were selected from each.

Selecting the Churches

The diocese in question publishes an annual report that lists the 100 or so diocesan churches and their membership sizes. This listing constituted the sampling frame for the first stage of sampling.

A total of approximately 500 respondents was desired for the study, so the decision was made to select 25 churches with probability proportionate to size and take 20 women from each of those selected. To accomplish this, the list of churches was arranged geographically, and then a table was created similar to the partial listing shown in Table 8-1.

Table 8-1 Form Used in Listing of Churches

Church	Membership	Cumulative Membership
Church A	3,000	3,000
Church B	5,000	8,000
Church C	1,000	9,000

Beside each church in the table, its membership was entered, and that figure was used to compute the cumulative total running through the list. The final total came to approximately 200,000. The object at this point was to select a sample of 25 churches in such a way that each would have a chance of selection proportionate to the number of its members. To accomplish this, the cumulative totals were used to create ranges of numbers for each church equaling the number of members in that church. Church A in the table was assigned the numbers 1 through 3,000; Church B was assigned 3,001 through 8,000; Church C was assigned 8,001 through 9,000; and so forth.

By selecting 25 numbers ranging between 1 and 200,000, it was possible to select 25 churches for the study. The 25 numbers were selected in a systematic sample as follows. The sampling interval was set at 8,000 (200,000/25), and a random start was selected between 1 and 8,000. Let us say the random number was 4,538. Because that number fell within the range of numbers assigned to Church B (3,001–8,000), Church B was selected.

Increments of 8,000 (the sampling interval) were then added to the random start, and every church within whose range one of the resultant numbers appeared was selected into the sample of churches. It should be apparent that in this fashion, each church in the diocese had a chance of selection directly proportionate to its membership size. A church with 4,000 members had twice the chance of selection as a church of 2,000 and 10 times the chance of selection as one with only 400 members.

Selecting the Churchwomen

Once the sample of churches was selected, arrangements were made to get lists of the women members of each. It is worth noting here that in practice the lists varied greatly in their form and content. In some cases, lists of all members (men and women) were provided, and it was necessary to sort out the women before sampling the lists. The form of the lists varied from typed lists to 3 × 5 cards printed from mailing address plates.

As the list arrived from a selected church, a sampling interval for that church was computed on the basis of the number of women members and the number desired (20). If a church had 2,000 women members, the sample interval, therefore, was set at 100. A random number was selected and incremented by the sampling interval to select the sample of women from that church. This procedure was repeated for each church.

Note that this sample design ultimately gives every woman in the diocese an equal chance of selection *only* if the assumption is made that each church has the same ratio of men to women. That is due to the fact that churches were given a chance of selection based on their

remainder of the city. Households in the suburban area, then, are given a disproportionately better chance of selection than those located elsewhere in the city.

As long as you analyze the two area samples separately or comparatively, you need not worry about the differential sampling. If you want to combine the two samples to create a composite picture of the entire city, however, you must take the disproportionate sampling into account. If n is the number of households selected from each area, then the households in the suburban area had a chance of selection equal to n divided by one-fourth of the total city population. Because the total city population and the sample size are the same for both areas, the suburban-area households should be given a weight of $\frac{1}{4}n$, and the remaining households should be given a weight of $\frac{3}{4}n$. This weighting procedure could be simplified by merely giving a weight of 3 to each of the households selected outside the suburban area. (This procedure gives a proportionate representation to each sample element. The population figure would have to be included in the weighting if population estimates were desired.)

Here's an example of the problems that can be created when disproportionate sampling is not accompanied by a weighting scheme. When the *Harvard Business Review* decided to survey its subscribers on the issue of sexual harassment at work, for example, it seemed appropriate to oversample women. Here's how Collins and Blodgett explained the matter:

> We also skewed the sample another way: to ensure a representative response from women, we mailed a questionnaire to virtually every female subscriber, for a male/female ratio of 68% to 32%. This bias resulted in a response to 52% male and 44% female (and 4% who gave no indication of gender)—compared to HBR's U.S. subscriber proportion of 93% male and 7% female.
>
> (1981:78)

You should have noticed a couple of things in this quotation. First, it would be nice to know a little more about what "virtually every female" means. Evidently, they didn't send questionnaires to all female subscribers, but there's no indication of who was omitted and why. Second, they didn't use the term *representative* in its normal social science usage. What they mean, of course, is that they want to get a substantial or "large enough" response from women, and oversampling is a perfectly acceptable way of accomplishing that.

By sampling more women than a straightforward probability sample would have produced, they have gotten enough women (812) to compare with the men (960). Thus, when the authors report, for example, that 32 percent of the women and 66 percent of the men agree that "the amount of sexual harassment at work is greatly exaggerated," we know that the female response is based on a substantial number of cases. That's good. There are problems, however.

To begin, subscriber surveys are always problematic. In this case, the best the researchers can hope to talk about is "what subscribers to *Harvard Business Review* think." In a loose way, it might make sense to think of that population as representing the more sophisticated portion of corporate management. Unfortunately, the overall response rate was 25 percent. Although that is quite good for subscriber surveys, it is a low response rate in terms of generalizing from probability samples.

Beyond that, however, the disproportionate sample design creates a further problem. When the authors state that 73 percent favor company policies against harassment (Collins and Blodgett, 1981:78), that figure is undoubtedly too high, since the sample contains a disproportionately high percentage of women—who are more likely to favor such policies. And, when the researchers report that top managers are more likely to feel that claims of sexual harassment are exaggerated than are middle- and lower-level management (1981:81), that finding is also suspect. As the researchers report, women are disproportionately represented in lower management. That alone might account for the apparent differences in different levels

of management. In short, the failure to take account of the oversampling of women confounds all survey results that don't separate findings by sex.

Degrees of Precision in Weighting

In any complex sample design, you face a number of options with regard to weighting of purposively or inadvertently disproportionate samples. You may compute weights for each element to several decimal places, or you may assign rough weights to account for only the grossest instances of disproportionate sampling. In the previous case of the city in which the suburban area was oversampled, it is unlikely that the population of that area was exactly one-fourth of the city's population: Suppose it actually was .25001, .2600, or .2816 of the total population. In each of these instances, whether you choose to apply the rough overall weighting of one-fourth or decide to be more precise will depend on the precision you desire in your findings. If your research purposes can tolerate errors of a few percentage points, you will probably not waste your time and effort in weighting exactly. In deciding the degree of precision required, moreover, you should take into account the degree of error to be expected from normal sampling distribution, plus all the various types of nonsampling error.

Ultimately, there is no firm guideline for you to follow in determining the precision to be sought in weighting. As in so many other aspects of study design, you have considerable latitude. At the same time, however, you should bear your decision in mind when reporting your findings. You should not employ only a rough weighting procedure and then suggest that your findings are accurate within a minuscule range of error.

Methods for Weighting

Having outlined the scientific concerns for determining the degree of precision desired in weighting, I should note that the choice will often be made on the basis of available methods for weighting. There are two basic methods for weighting.

1. For the rough weighting of samples drawn from subpopulations, weighted tables can be constructed from the unweighted tables for each of the subsamples. In the earlier example, you could create a raw table of distributions for the suburban sample and for the nonsuburban sample separately, triple the number of cases in each cell of the nonsuburban table, add the cases across the two tables, and compute percentages for the composite table.

2. If the data are to be analyzed by computer, a special program may be designed to assign a precise weight to each case in the data file. This way, the computer does all the work of weighting.

It bears repeating that complex weighting methods are usually not necessary in social research. Understanding the logic of weighting, however, will strengthen your grasp of the logic of sampling in general.

Figure 8-13 offers a graphic illustration of the logic of weighting. You'll recall that our micropopulation of 100 people (Figure 8-1) contains only 12 African Americans—6 men and 6 women (which is the approximate percentage of African Americans in America). Representative samples from such a population would provide few African-American subjects for study, which is a problem if we wished to compare African Americans and whites, for example. We might resolve this problem by oversampling African Americans, then weighting the resulting data.

In the illustration, we've selected all the African Americans into our sample, but only one-fourth of the whites. Having collected the data, we've then weighted each white respondent's data by a factor of four. If we did this by actually making copies of the data files, we would end up with 100 data files—which could be analyzed as though we had collected data from all 100 members of the population.

Figure 8-13 Sampling Disproportionately and Weighting

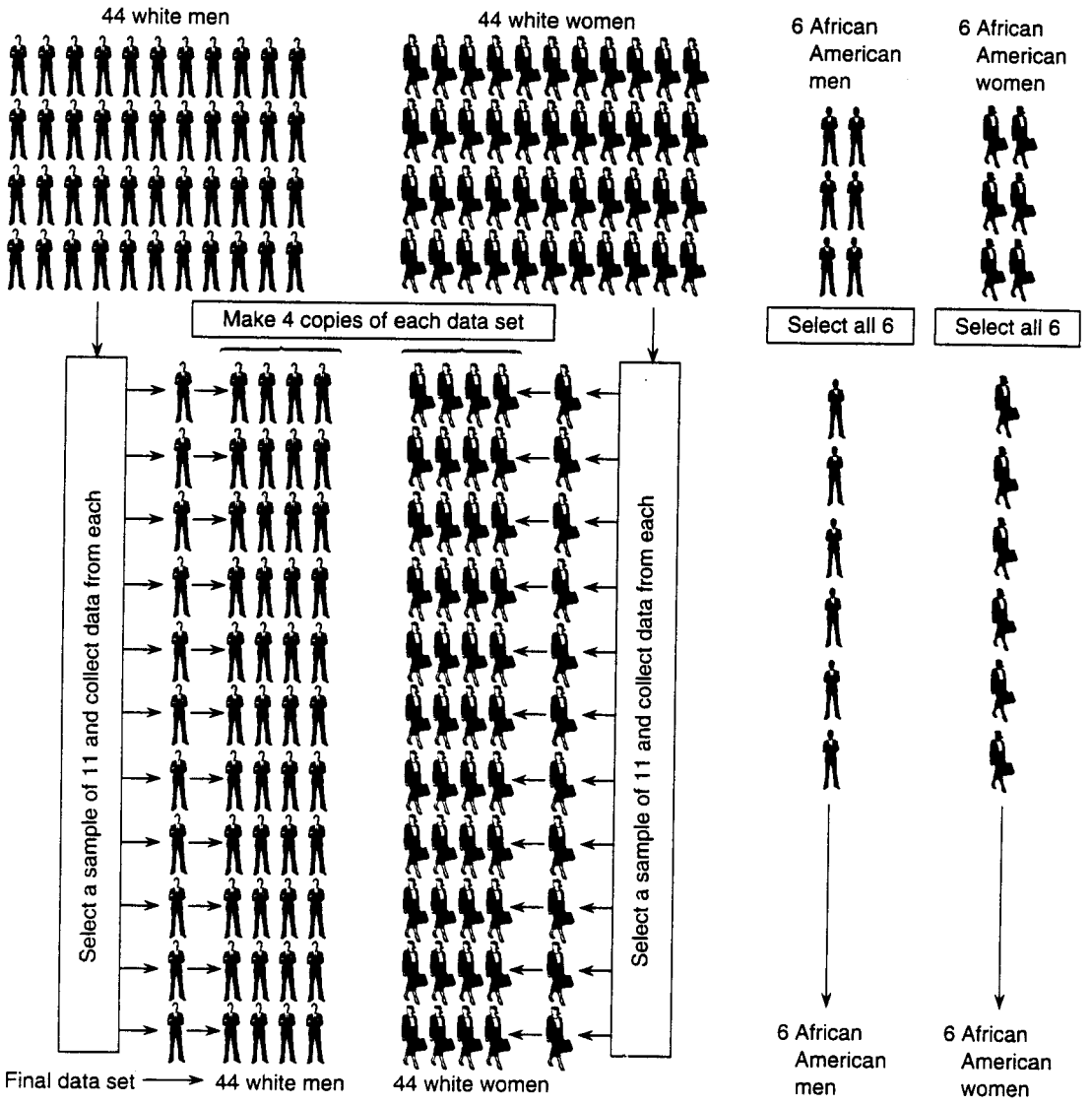


Illustration: Sampling Churchwomen

Let's complete the discussion of cluster sampling with a research example. The illustration that follows is not as complex as the **area probability samples** that are employed in studies of geographic areas such as cities, states, or the nation. Nonetheless, it should be a useful example of the various principles of cluster sampling.

The purpose of this study was to examine the religious and social attitudes of women members of churches in a diocese of the Episcopal church. A representative sample of all churchwomen in the diocese was desired. Because there was no single list of such women, a multi-

stage sample design was created. In the initial stage of sampling, churches were selected with probability proportionate to size (PPS), and then women were selected from each.

Selecting the Churches

The diocese in question publishes an annual report that lists the 100 or so diocesan churches and their membership sizes. This listing constituted the sampling frame for the first stage of sampling.

A total of approximately 500 respondents was desired for the study, so the decision was made to select 25 churches with probability proportionate to size and take 20 women from each of those selected. To accomplish this, the list of churches was arranged geographically, and then a table was created similar to the partial listing shown in Table 8-1.

Table 8-1 Form Used in Listing of Churches

Church	Membership	Cumulative Membership
Church A	3,000	3,000
Church B	5,000	8,000
Church C	1,000	9,000

Beside each church in the table, its membership was entered, and that figure was used to compute the cumulative total running through the list. The final total came to approximately 200,000. The object at this point was to select a sample of 25 churches in such a way that each would have a chance of selection proportionate to the number of its members. To accomplish this, the cumulative totals were used to create ranges of numbers for each church equaling the number of members in that church. Church A in the table was assigned the numbers 1 through 3,000; Church B was assigned 3,001 through 8,000; Church C was assigned 8,001 through 9,000; and so forth.

By selecting 25 numbers ranging between 1 and 200,000, it was possible to select 25 churches for the study. The 25 numbers were selected in a systematic sample as follows. The sampling interval was set at 8,000 (200,000/25), and a random start was selected between 1 and 8,000. Let us say the random number was 4,538. Because that number fell within the range of numbers assigned to Church B (3,001–8,000), Church B was selected.

Increments of 8,000 (the sampling interval) were then added to the random start, and every church within whose range one of the resultant numbers appeared was selected into the sample of churches. It should be apparent that in this fashion, each church in the diocese had a chance of selection directly proportionate to its membership size. A church with 4,000 members had twice the chance of selection as a church of 2,000 and 10 times the chance of selection as one with only 400 members.

Selecting the Churchwomen

Once the sample of churches was selected, arrangements were made to get lists of the women members of each. It is worth noting here that in practice the lists varied greatly in their form and content. In some cases, lists of all members (men and women) were provided, and it was necessary to sort out the women before sampling the lists. The form of the lists varied from typed lists to 3 × 5 cards printed from mailing address plates.

As the list arrived from a selected church, a sampling interval for that church was computed on the basis of the number of women members and the number desired (20). If a church had 2,000 women members, the sample interval, therefore, was set at 100. A random number was selected and incremented by the sampling interval to select the sample of women from that church. This procedure was repeated for each church.

Note that this sample design ultimately gives every woman in the diocese an equal chance of selection *only* if the assumption is made that each church has the same ratio of men to women. That is due to the fact that churches were given a chance of selection based on their

total membership with an assumption that half the members of each church were women. In fact, of course, women made up more than half the members of some churches and less than half of others. Given the aims of this particular study, the slight inequities of selection were considered insignificant.

A more sophisticated sample design for the second stage would have resolved this possible problem. Because each church was given a chance of selection based on an assumed number of women (assuming 1,000 women in a church of 2,000), the sampling interval could have been computed on the basis of that assumption rather than on the actual number of women listed. If it were assumed in the first stage of sampling that a church had 1,000 women (out of a membership of 2,000), the sampling interval could have been set at 50 (1,000/20). Then this interval could have been used in the selection of respondents regardless of the actual number of women listed for that church. If 1,000 women were in fact listed, then their church had the proper chance of selection and 20 women would be selected from it. If 1,200 women were listed, that would mean that the church had too small a chance of selection, but this would have been remedied through the selection of 24 women using the preestablished sampling interval. If only 800 women were listed, on the other hand, only 16 would have been selected.

Probability Sampling in Review

The preceding discussions have been devoted to the key sampling method used in controlled survey research: probability sampling. In each of the variations examined, we have seen that elements are chosen for study from a population on a basis of random selection with known nonzero probabilities.

Depending on the field situation, probability sampling can be very simple, or it can be extremely difficult, time consuming, and expensive. Whatever the situation, however, it remains the most effective method for the selection of study elements. There are two reasons for this.

First, probability sampling avoids conscious or unconscious biases in element selection on the part of the researcher. If all elements in the population have an equal (or unequal and subsequently weighted) chance of selection, there is an excellent chance that the sample so selected will closely represent the population of all elements.

Second, probability sampling permits estimates of sampling error. Although no probability sample will be perfectly representative in all respects, controlled selection methods permit the researcher to estimate the degree of expected error in that regard.

In spite of the preceding comments, it is sometimes not possible to use standard probability sampling methods. Sometimes it isn't even appropriate to do so. In those cases, nonprobability sampling is used. The remainder of this chapter is devoted to a brief discussion of the different forms of nonprobability sampling available to you as a researcher.

Nonprobability Sampling

I'm sure you can envision situations in which it would be either impossible or unfeasible to select the kinds of probability samples I described. Suppose you wanted to study embezzlers: There is no list of all embezzlers nor are you likely to create such a list. Moreover, as we'll see, there are times when probability sampling wouldn't be appropriate if possible. In many such situations, nonprobability sampling procedures are called for, and we'll examine three types in this section: purposive or

judgmental sampling, quota sampling, and the reliance on available subjects.

Purposive or Judgmental Sampling

Occasionally it may be appropriate for you to select your sample on the basis of your own knowledge of the population, its elements, and the nature of your research aims: in short, based on your judgment and the purpose of the study. Especially in the initial design of your questionnaire, you might wish to select the widest variety of respondents to test the broad applicability of questions. Although the study findings would not represent any meaningful population, the test run might effectively uncover any peculiar defects in your questionnaire. This situation would be considered a pretest, however, rather than a final study.

In some instances, you may wish to study a small subset of a larger population in which many members of the subset are easily identified, but the enumeration of all of them would be nearly impossible. For example, you might want to study the leadership of a student protest movement; many of the leaders are easily visible, but it would not be feasible to define and sample all leaders. In studying all or a sample of the most visible leaders, you may collect data sufficient for your purposes.

In a multistage sample design, you might want to compare left-wing and right-wing students. Because you may not be able to enumerate and sample from all such students, you might decide to sample the memberships of the Green Party and Young Americans for Freedom. Although such a sample design would not provide a good description of either left-wing or right-wing students as a whole, it might suffice for general comparative purposes.

Sampling of *selected precincts* for political polls is a somewhat refined judgmental process. On the basis of previous voting results in a given area (city, state, nation), you purposively select a group of voting precincts that, in combination, produces results similar to those of the entire area. Then, in later polls, you select your samples solely from those precincts. The theory is that the selected precincts provide a cross section of the entire electorate, so you need to know what you're doing.

Each time an election is held, permitting you to evaluate the adequacy of your group of precincts, you would consider revisions, additions, or deletions. Your goal is to update the group of precincts to ensure that it will provide a good representation of all precincts.

To be done effectively, selected precinct sampling requires considerable political expertise. You should be well versed in the political and social history of the area under consideration so that the selection of precincts is based on an *educated* guess about its persistent representativeness. In addition, this system of sampling requires continuing feedback to be effective. You must be in a position to conduct frequent polls and must have periodic electoral validations.

Quota Sampling

Quota sampling, mentioned earlier, is the method that helped George Gallup avoid disaster in 1936—and set up the disaster of 1948. Like probability sampling, quota sampling addresses the issue of representativeness, though the two methods approach the issue quite differently.

Quota sampling begins with a matrix describing the characteristics of the target population. You need to know what proportion of the population is male and what proportion female, for example, and for each sex, what proportions fall into various age categories, educational levels, ethnic groups, and so forth. In establishing a national quota sample, you would need to know what proportion of the national population is urban, eastern, male, under 25, white, working class, and the like, and all the other permutations of such a matrix.

Once such a matrix has been created and a relative proportion assigned to each cell in the matrix, you collect data from persons having all

the characteristics of a given cell. All the persons in a given cell are then assigned a weight appropriate to their portion of the total population. When all the sample elements are so weighted, the overall data should provide a reasonable representation of the total population.

Quota sampling has several inherent problems. First, the quota frame (the proportions that different cells represent) must be accurate, and it is often difficult to get up-to-date information for this purpose. The Gallup failure to predict Truman as the presidential victor in 1948 was due partly to this problem.

Second, biases may exist in the selection of sample elements within a given cell—even though its proportion of the population is accurately estimated. An interviewer, instructed to interview five persons meeting a given, complex set of characteristics, may still avoid persons living at the top of seven-story walk-ups, having particularly run-down homes, or owning vicious dogs.

In recent years, attempts have been made to combine probability and quota sampling methods, but the effectiveness of this effort remains to be seen. At present, it is a good idea to treat quota sampling warily.

Reliance on Available Subjects

Relying on available subjects, that is, stopping people at a street corner or some other location, is almost never an adequate sampling method, although it is used all too frequently. It is justified only if the researcher wants to study the characteristics of people passing the sampling point at specified times.

University researchers frequently conduct surveys among the students enrolled in large lecture classes. The ease and inexpense of such a method explains its popularity, but it seldom produces data of any general value. It may be useful to pretest a questionnaire, but such a sampling method should not be used for a study purportedly describing students as a whole.

Consider this report on the sampling design in an examination of knowledge and opinions about nutrition and cancer among medical students and family physicians:

> The fourth-year medical students of the University of Minnesota Medical School in Minneapolis comprised the student population in this study. The physician population consisted of all physicians attending a "Family Practice Review and Update" course sponsored by the University of Minnesota Department of Continuing Medical Education.
>
> (Cooper-Stephenson and Theologides, 1981:472)

After all is said and done, what will the results of this study represent? They do not provide a meaningful comparison of medical students and family physicians in the United States or even in Minnesota. Who were the physicians who attended the course? We can guess that they were probably more concerned about their continuing education than other physicians, but we can't say that for sure. Ultimately, we don't know what to do with the results of a study like this.

Main Points

- A sample is a special subset of a population observed for purposes of making inferences about the nature of the total population itself.
- The sampling methods used earlier in this century often produced misleading inferences, and current techniques are far more accurate and reliable.
- The chief criterion of the quality of a sample is the degree to which it is representative—the extent to which the characteristics of the sample are the same as those of the population from which it was selected.
- Probability sampling methods provide one excellent way of selecting samples that will be quite representative.
- The most carefully selected sample will almost never provide a perfect representation