

4

Equal Probability of Selection

Source: Frerichs, R.R. Rapid Surveys (unpublished), © 2004. NOT FOR COMMERCIAL DISTRIBUTION

4.1 INTRODUCTION

When is a person treated as a person? For rapid surveys, the answer is when the sample is *self-weighted*. Statistical calculations are easier when each person in a population has the same chance of being selected as any other person. On occasion, we may draw a sample that is not self-weighted. When this occurs, we need to add numeric weights to our estimates of the average value and standard error, so that the sample again gives a factual view of the population. The use of such weights will be described in Chapter 6. For now, however, I will focus on self-weighted sampling.

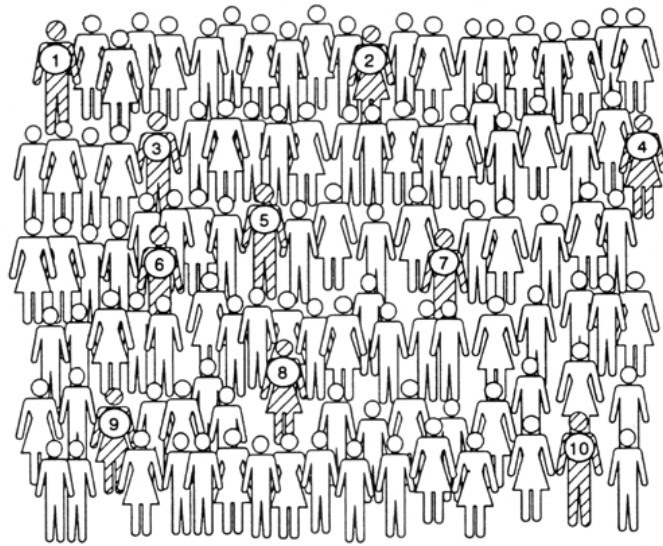


Figure 4-1. Random sample of 10 persons from a population of 120 persons.

4.1.1 Sample of ten from population of 120

In Chapter 2 we learned of simple random sampling with a population of nine addicts and samples of three addicts. We observed that with such a small population there are only 84 possible samples that could be drawn of three addicts without replacement. I concluded the chapter with an example of a survey of smoking patterns, with a sample of 30 households from 1,000 households and a sample of 90 persons from a population of 3,000 persons. We discovered that the number of possible samples is greater than we could easily count, or even that a computer could count. As a result, we had to draw samples from a huge list of all possible samples so as to create and view 100 confidence intervals. Here I will present still another sample but this time of ten people from a population of 120. I will use this example to illustrate what it means to be a *self-weighted* sample and for each person to have the same probability of being selected.

Assume there is population of 120 persons from which we are drawing a random sample of 10 persons (see Figure 4-1). All of the persons were sampled *without replacement* so that the 10 persons in the sample are all different from one another. There are many possible combinations of 10 people that could be selected from the population. Using Formula 3.2, we can calculate the exact number of possible samples as:

$$\frac{120!}{10! (120-10)!}$$

Extending the formula, the calculations are...

$$\frac{120 \times 119 \times 118 \times 117 \times 116 \times 115 \times 114 \times 113 \times 112 \times 111 \times 110!}{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 \times 110!}$$

The term *110!* cancels in the numerator and denominator,

$$\frac{120 \times 119 \times 118 \times 117 \times 116 \times 115 \times 114 \times 113 \times 112 \times 111}{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}$$

resulting in an answer of 116,068,178,638,776 possible samples. If we had much paper and endless time, we could list all 116 trillion possible samples, as represented in Figure 4-2. Keep in mind that the one sample we actually selected (shown in the lower left of Figure 4-2) is merely one of the 116 trillion samples we could have selected. While our survey of 10 persons may be viewed as a random sample of 10 individuals from a list of 120 people, you could also consider it a random sample of a group of 10 persons from a list of 116 trillion possible samples. If the sampling method is self-weighted and free of bias, each of these 116 trillion samples has the same probability of being selected. By chance alone, we selected sample number 100,000,000,000,000.

The data in each of the possible samples can be used to derive a proportion or a mean, depending on whether the variable is binomial or equal interval (see Figure 4-3). The mean and proportion of the sample we selected (number 100,000,000,000,000) is shown in the lower left section of Figure 4-3, listed among the 116,068,178,638,775 other possible samples. If the sampling method is self-weighted and unbiased, each of the proportions or means from the trillions of samples will have

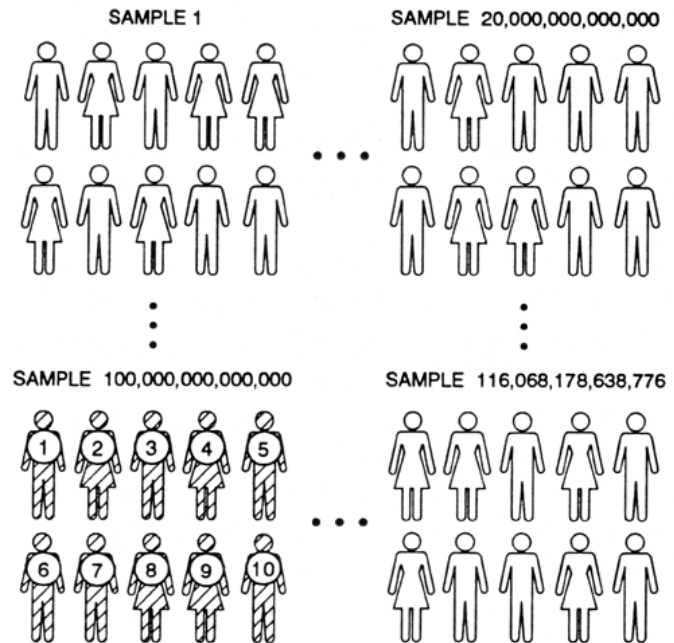


Figure 4-2. Possible samples of 10 persons from a population of 120, drawn without replacement.

the same probability of being selected and we will be able to use the statistical formulae cited in Chapter 3 (and presented in more detail in Chapter 5) to calculate the standard error and confidence interval. If the sample is biased, some of the possible samples may have no chance of being selected. If the persons are not self-weighted, some of the possible samples will be more likely to be selected than others. Different from some forms of bias, however, we can correct for unequal probabilities of selection with numeric weights that we attach to each sampled person. Such weighting is common in advanced surveys but is not necessary with most types of rapid survey. I will, however, describe weights in Chapter 6, as a useful procedure for combining rapid surveys from several regions of a country, or survey strata of different groups identified by race, sex, education and the like.

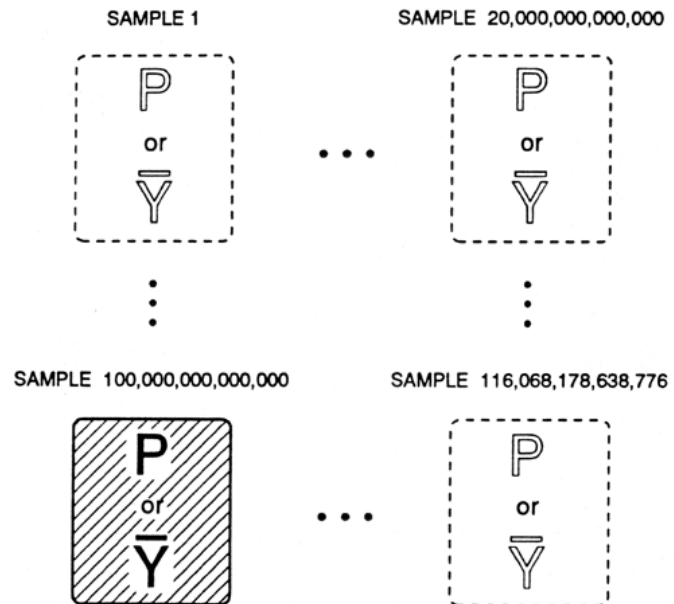


Figure 4-3. Proportion or mean of possible samples of 10 persons from a population of 120 persons drawn without replacement.

4.2 EPSEM SAMPLING

Sampling which results in each person having the same chance of being selected is termed *equal probability of selection method* (EPSEM) sampling. For sampling done in one stage, such as simple random sampling described in Chapter 2, each person should have the same probability of being selected at the beginning of the stage. If the sample is drawn in two stages, as will be described in this chapter, the probability of being selected may be different at the beginning of stage one than in advance of stage two. EPSEM sampling does not mean that every unit at every stage has the same probability of being selected. Instead it means that every unit that is actually included in the sample had the same probability of being

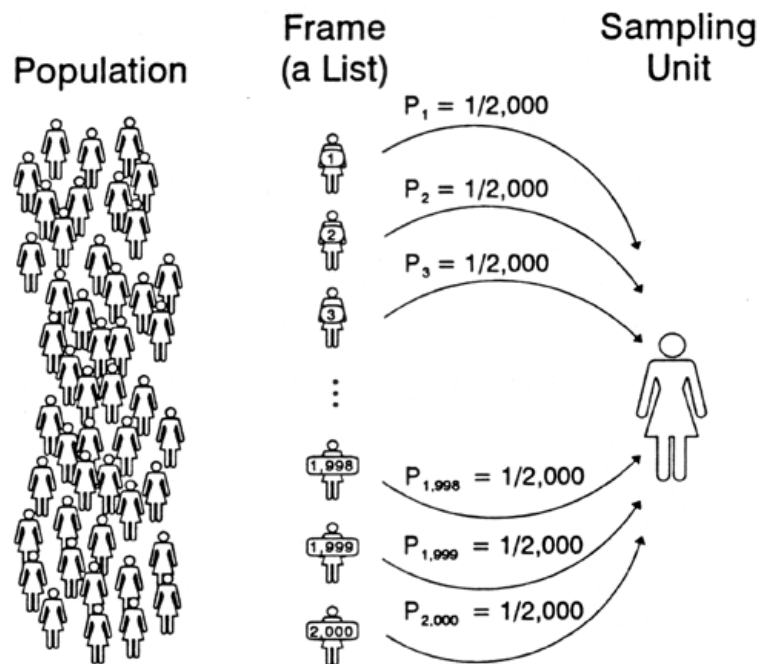


Figure 4-4. Equal probability of selection of a person in a population of people.

selected in advance of sampling. Noone is left out and noone is favored.

4.2.1 One-stage sampling

Typical of EPSEM surveys is a simple random sample of women selected in one stage from a population of 2,000. As observed in Figure 4-4, 2,000 women comprise a population. That is, the *population* is people. Before being surveyed, they must all be included in a larger list termed a *frame*, with sequential identification numbers starting at one and ending with the total number in the population. In our example the identification numbers go from 1 to 2,000. Using random numbers between 1 and 2,000, ten women are sampled from the frame. Thus the *sampling unit* is a person. If the sample is self-weighted, each woman has a probability of being selected of 1/2,000 or 0.0005.

The concept is the same when sampling households (see Figure 4-5). Rather than a population of persons, Figure 4-5 shows a *population* of 1,000 households. To do a simple one-stage random sample, each household is included in a *frame* with a sequential identification number from one to 1,000. If the sample is self-weighted, each household has the same probability of being selected, shown in Figure 4-5 as 1/1,000 or 0.001. Notice that the *sampling unit* is households, not people and that the sampled households are selected in one stage from the frame.

The probability of being selected for a sample increases with the number of units that are sampled. Assume that you are sampling 20 units (either people or households) from a population of 10,000 units. Each person has a probability of 1/10,000 or 0.0001.

From basic probability theory we know that for a sample of two units, the probability that a unit is selected at either drawing one or two is defined as in Formula 4.1

$$P(\text{one or two}) = P(\text{one}) + P(\text{two}) \quad (4.1)$$

The selection probability for a sampling unit when there is multiple drawings is the addition of the individual probabilities. Returning to our example of a sample of 20 units from a population of 10,000 units, the selection probability for entering the sample is...

$$P(\text{one or two or... 19 or 20}) = P(\text{one}) + P(\text{two}) + \dots P(19) + P(20)$$

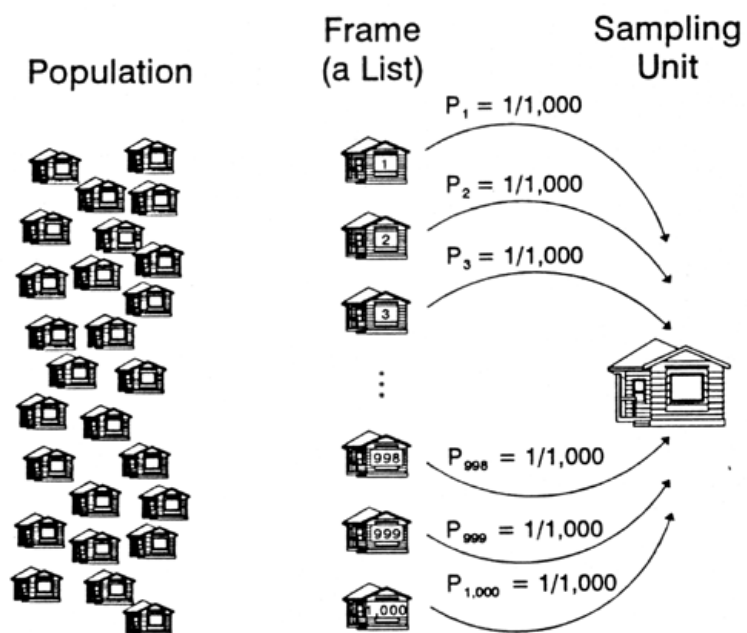


Figure 4-5. Equal probability of selection of a household in a population of households.

or 20/10,000, as shown in Figure 4-6. To be unbiased, the probability of any one unit being selected should be the same as any other unit. That is, the selection probability for each unit should be 20/10,000 or 0.002. If the selection probabilities are different, the sample units can be mathematically weighted so that each sample unit again represents the same number of units in the population. Such weighting schemes, however, are more complicated and should be avoided when doing rapid surveys. The concept of *weighting* will be presented further in Chapter 6.

4.2.2 Two-stage sampling

Persons or households can also be selected in two stages. The first stage might be a sample of city blocks or districts, or villages in rural areas. The second stage is a sample of people or households drawn from those units selected in the first stage. An example of the first stage selection process is shown in Figure 4-7 for a small town with 65 blocks. The blocks are listed with sequential numbers from 1 to 65. At the first stage, a sample of nine blocks is drawn from the population of 65 blocks. Notice that time and effort must be spent in listing the blocks at the first stage. There is no easy way to do this step other than to count the blocks.

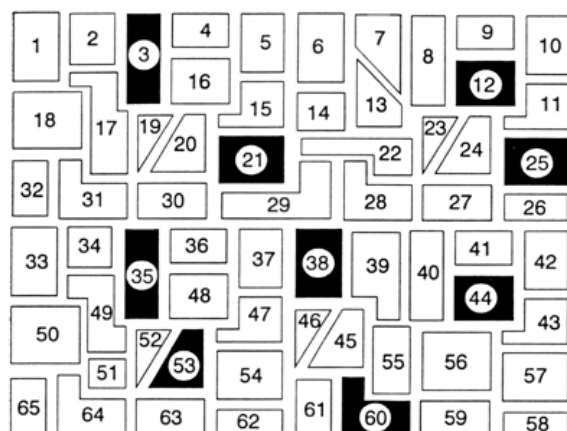


Figure 4-7. Listing of blocks in sampling frame and selection of blocks at first stage of two-stage sampling.

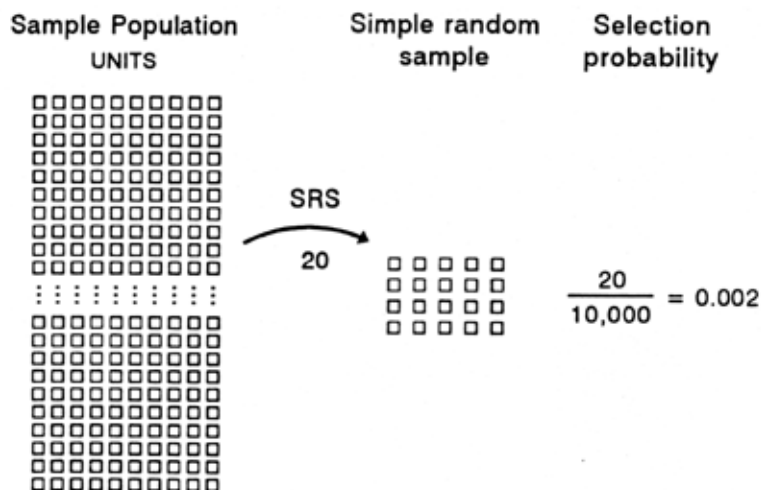


Figure 4-6. Simple random sample of 20 units from a population of 10,000 units.

Only those blocks that are selected in Figure 4-7 at the first stage are included in the second stage of the sampling process. Yet before further sampling can be done, we again need to create a frame (or list) of persons or households in each sampled block. That is, a sampling frame must be created for blocks number 3, 12, 21, 25, 35, 38, 44, 53 and 60, but not for the remaining 56 blocks. If we had done a one-stage simple random sample, we would have to have listed every *sampling unit* in the population rather than just those in the nine selected blocks. This would have been much more time consuming and is the main reason why simple random sampling is rarely done. The field work is too costly, unless of course there are volunteers available who will do

quality work for little or no pay. For most people, however, two-stage sampling is preferable.

For the second stage, only those blocks to be sampled are included. Thus in our example, a new frame is created with nine blocks (see Figure 4-8). Each of the nine blocks did not have the same chance of being selected. Instead, the blocks at the first stage were selected with probability proportionate to their size (PPS). Using this method, larger blocks are more likely to be included in our sample than smaller blocks. Further details on this selection procedure, featured in rapid surveys, are presented in Section 4.3.

Within the nine selected blocks, all persons or households are listed. A simple random sample is then done in each cluster of the listed units and the data are combined into a single estimate for the survey as a whole. Figure 4-9 illustrates the two-stage sampling process for a survey of women to assess their choice of contraceptive methods for family planning. The 65 city blocks are listed in the first stage *frame*. Nine of the blocks are sampled from the frame with probability proportionate to size. These nine blocks lead to the creation of another frame, consisting of nine lists of eligible women living in the selected blocks. Once each list is made, seven women are selected from the list by simple random sampling. The total sample consists of 63 women, too small for a rapid survey but large enough for an example.

Rather than persons, the second stage of the sampling process is often households (see Figure 4-10). The sampling process is identical except that the second stage *frame* consists of a listing of households in each block rather than women. The size of the second survey is 36 households, far smaller than the first survey. But is it? As you will learn later, in household surveys we are generally not interested in characteristics of the household. Instead we want to know more about the occupants of the households. If we collected information on each persons in the 36 households, and there was an average of four persons per household, the survey would consist of 144 persons, compared to 63

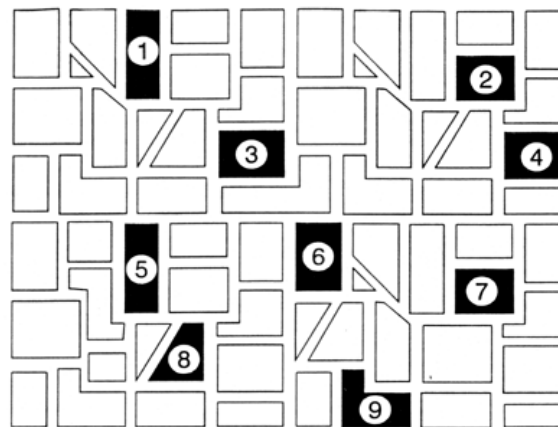


Figure 4-8. Listing of population for second stage sampling only in blocks selected at first stage.

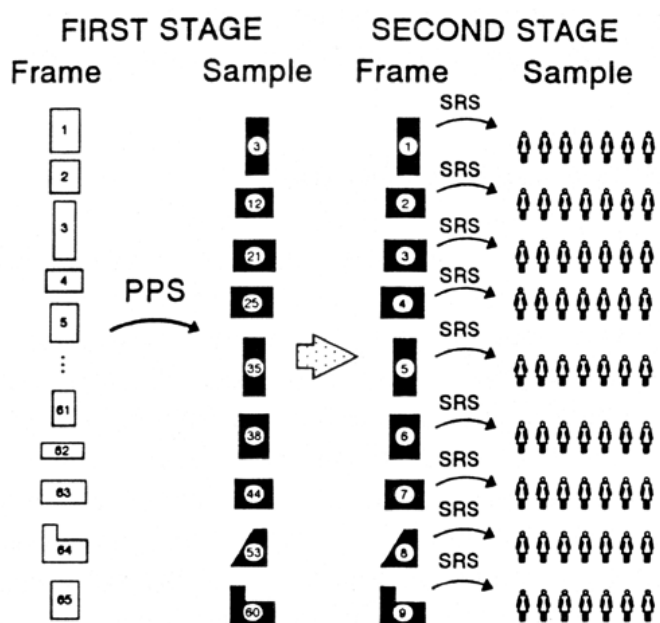


Figure 4-9. Two stage sampling with blocks sampled at first stage with probability proportionate to size and persons sampled at second stage with simple random sampling.

with the family planning survey of women.

Conversely, if we only wanted information on women, aged 20 through 44 years, in the households, the number of participants in our sampled households might be much smaller than 63.

4.2.3 Two-stage cluster sampling

While the concepts of *self-weighting* and *equal probability of selection* are easy to read about, they are sometimes overlooked when planning a survey. In this section I will present four sampling strategies for two-stage cluster surveys, three of which are EPSEM samples and one which is not. Rather than city blocks, households or people, I will use circles and squares to represent sampling units. The population to be sampled usually has some organizational characteristics that links groups of people together at some time of day or year. The group may be defined as living in a common census tract or enumeration district, or in going to the same school, or being eligible to attend the same health clinic. The term *cluster* is used by sampling specialists to describe units in the population that contain the sampling units of interest. A cluster is formally defined as a natural grouping within a population. This grouping may be a city block, a census tract or enumeration district, a health district, a school district, or a community such as a village, hamlet, or town. For rapid surveys, clusters are usually geographic areas with political or social boundaries. The cluster may vary in size, although on rare occasions the clusters may all be the same size.

To illustrate how procedures at two steps of the sampling process can bias a survey, I will use the same example presented earlier in Figure 4-6. There we sampled 20 units from a population or 10,000 units and the selection probability was 0.002. We will again draw a sample of 20 from 10,000 but do so in two stages to ease the field work. If the sampling scheme is unbiased, each unit in the population should have the same probability of appearing in the sample. In our example, that selection probability is 0.002. I have divided the population into 200 *clusters* or groupings of units. In Figure 4-11, I show what occurs when the clusters vary in size, with some being as small as 10 units and others being as large as 100 units. At the first stage, four clusters are selected by simple random sampling. Thus each cluster has the same probability of being selected: $1/200$ or 0.005. At the second stage, I select an *equal fraction* of units from each cluster drawn at the first stage, again by simple random sampling. As a result, 10 units would be selected from the larger cluster of 100 units, while one unit would be drawn from the small cluster of 10 unit. How does this two-part selection scheme effect the selection

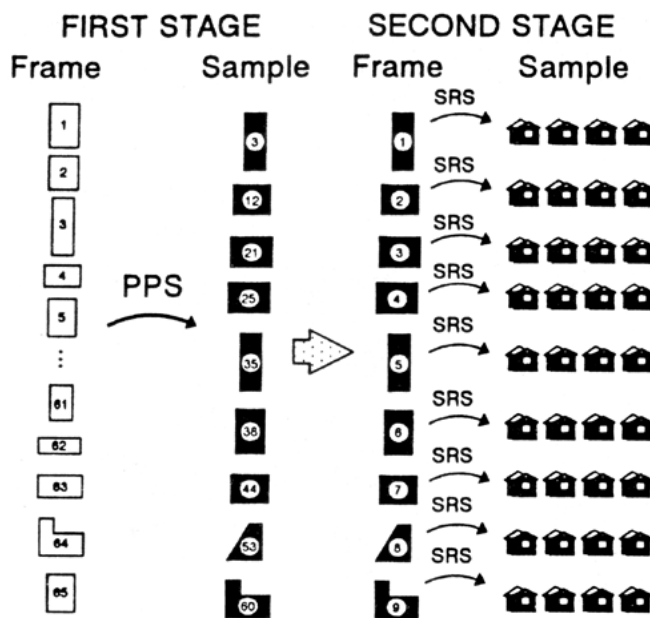


Figure 4-10. Two stage sampling with blocks sampled at first stage with probability proportionate to size and households sampled at second stage with simple random sampling.

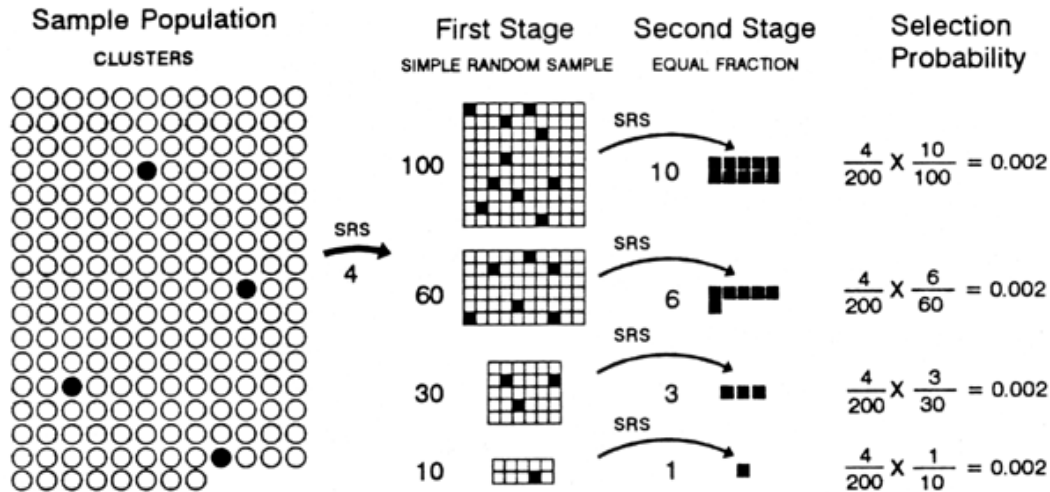


Figure 4-11. Equal probability of selection with simple random sampling of unequal sized-clusters at first stage and simple random sampling of equal fraction at second stage.

probabilities? To answer this question we need to again use probability theory. Basic probability theory states that for a sampling scheme with two stages, the probability of having a unit selected at both stages is defined as in Formula 4.2

$$P(\text{Stage one and Stage two}) = P(\text{Stage one}) \times P(\text{Stage two}) \quad (4.2)$$

If a unit happens to be in a cluster of 100 units, the probability of being selected with the described sampling scheme is 1/200 at the first stage and 10/100 at the at the second stage for a selection probability of...

$$P(\text{Stage one and Stage two}) = \frac{1}{200} \times \frac{10}{100}$$

or 0.002. This selection probability is the same for all 20 units included in the sample, as it was in Figure 4-6. Hence if clusters vary in size, an unbiased method of sampling is to draw clusters at the first stage by simple random sampling, followed by a simple random sample of a constant fraction of units in the selected clusters at the second stage.

For a second example, assume that all clusters are the same size (see Figure 4-12). Here we draw four clusters at the first stage by simple random sampling. At the second stage, however, we draw an *equal number* of units in the selected clusters, again by simple random sampling. The selection probabilities for all units in the sample remains at 0.002, indicating that this second sampling scheme also does not favor one unit over another.

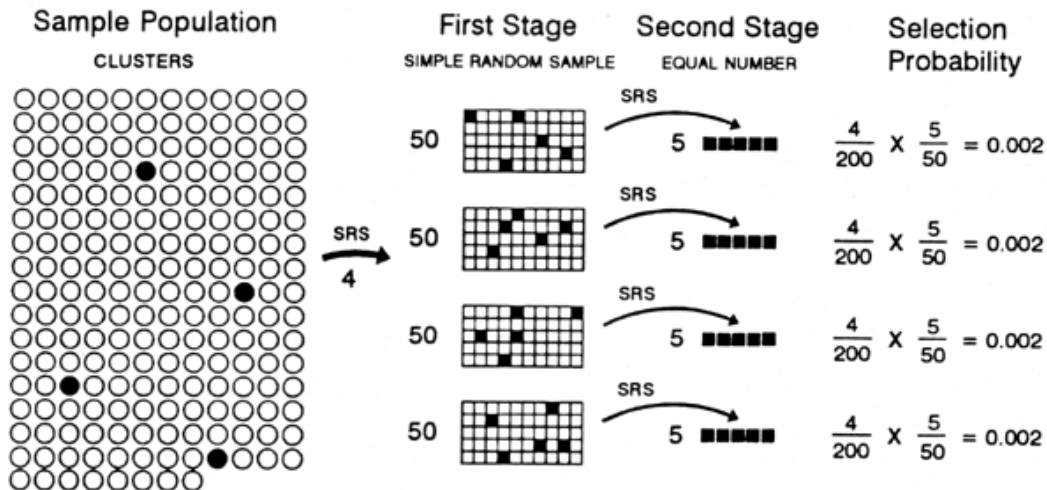


Figure 4-12. Equal probability of selection with simple random sampling of equal-sized clusters at first stage and simple random sampling of equal number at second stage.

The next sampling approach, shown in Figure 4-13, does not work so well. Here again we have clusters of varying size, typical of most field situations. Instead of sampling an equal fraction of units at the second stage, we select an equal number of units. Observe in Figure 4-13 that the

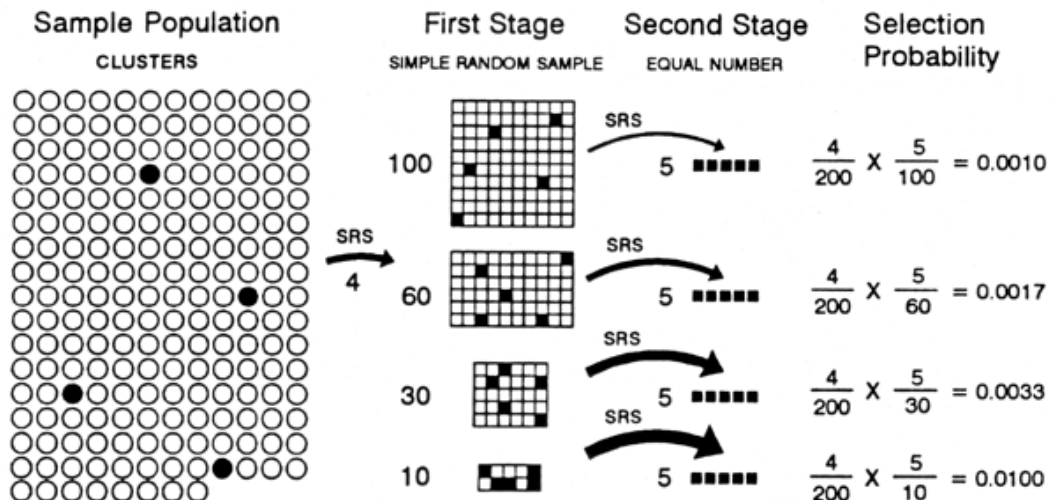


Figure 4-13. Unequal probability of selection with simple random sampling of unequal-sized clusters at first stage and simple random sampling of equal number at second stage.

selection probabilities vary greatly from a low of 0.001 in the largest clusters of 100 units to a high of 0.01 in the smallest clusters of 10 units. That is, units in small clusters are 10 times more likely to be sampled than units in larger clusters. Such a scheme might result in a biased sample, especially if units in large clusters are different in some important way from units in small clusters.

So far we have learned that when clusters vary in size and first stage selection is a simple random sample of clusters, an *equal fraction* of units must be selected at the second stage. Yet to save time in the field and make the procedure less complicated for field staff, we may not want to calculate a fraction to be sampled in each cluster. Instead we may want to assign the field staff to select an *equal number* in each cluster. The only way that we can do this is to either redraw the boundaries of clusters so that each has the same number of units, as shown in Figure 4-12, or to change the manner in which we sample clusters at the first stage. The latter is what we do for rapid surveys.

Figure 4-14 shows a sampling scheme which uses a different approach to select clusters at the first stage. Observe that the clusters again vary in size as they did in Figures 4-11 and 4-13. Also notice that the second stage selection is of an *equal number* of units, not an equal fraction. The clusters at the first stage are selected with probability proportionate to their size (PPS). That is, larger clusters are more likely to be drawn into a sample than smaller clusters. At first, this may seem like an unfair process since units in large clusters would be favored over units in small clusters. Keep in mind, however, that the selection process involves two stages. We should view the selection probabilities at the end of the selection process rather than just at the first stage. Returning

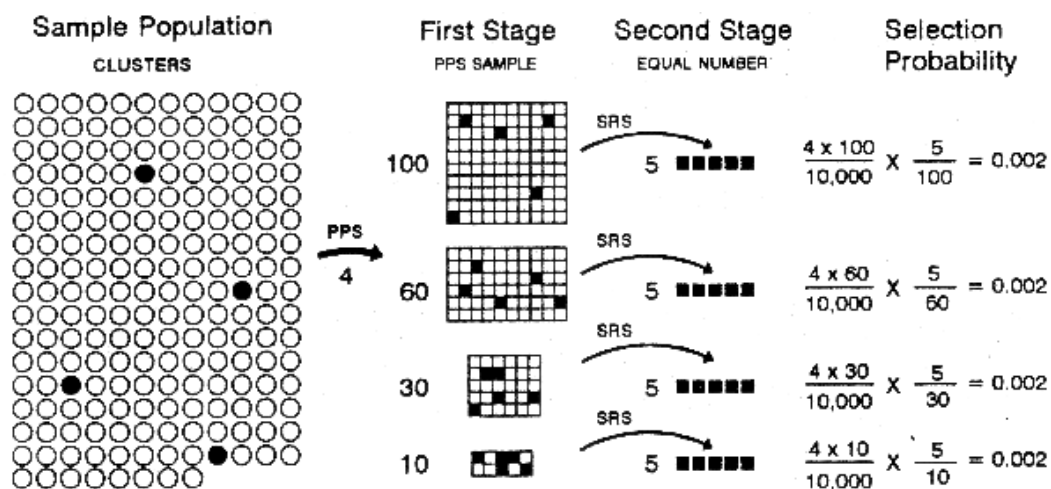


Figure 4-14. Equal probability of selection with probability proportionate to size sampling of unequal-sized clusters at first stage and simple random sampling of equal number at second stage.

to Figure 4-14, if the sampling unit is in a large cluster of 100, the probability of being selected is dependent on the number of units to be sampled (four) and the probability of being sampled. Since the probability of being sampled is proportionate to the size, for clusters of 100 units the probability is 100/10,000 or 0.01. Since four clusters are sampled, the probability of being selected in a cluster with 100 units is 4 times 0.01 or 0.04. The same logic holds for the other sized clusters with probabilities at the first stage of 0.024 for clusters of 60 units, 0.012 for clusters of 30 units and 0.004 for clusters of 10 units. When the second stage of sampling is completed, however, the selection probabilities are all 0.002, the same as in Figures 4-6, 4-11 and 4-12. Thus when clusters are unequal in size and you want to select at the second stage an *equal number* of unit per cluster, the first stage selection of clusters should be done with *probability proportionate to size* rather than

by simple random sampling.

4.3 PPS SAMPLING

Sampling with probability proportionate to size (PPS) sounds more complicated than it really is. To do PPS sampling of clusters, the various clusters in the population are assembled in a cumulative list as shown in Figure 4-15. The units in the list depend on what is to be selected at the second stage of the sampling process. If the persons are to be selected, the units in the cumulative list are people, while if households are to be drawn, the units are households. Each cluster has a position on the cumulative list.

Once the list is assembled, a random sample is taken of the list for each cluster to be visited. The procedure for the sampling process is shown in Figure 4-16. For example, assume that there are 200 units in the population. The first community has 100 units, the second has 60 units, the third has 30 units and the fourth has 10

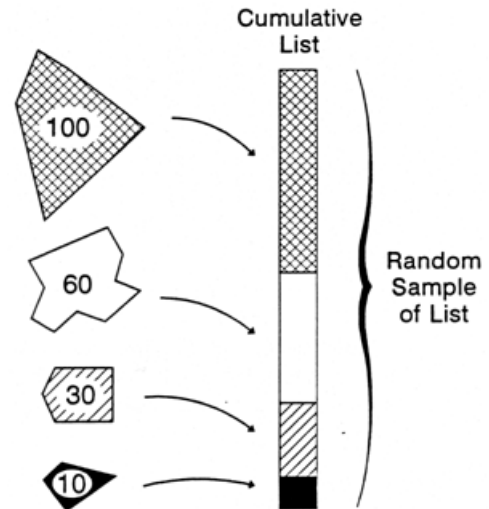


Figure 4-15. Cumulative listing of population for probability proportionate to size sampling.

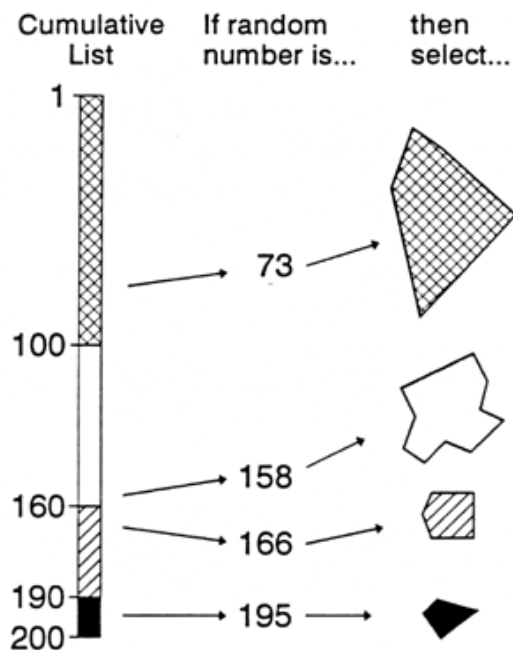


Figure 4-16. Probability proportionate to size sampling of clusters from a cumulative list of the population.

units. The first community consists of half the list and thus should have an easier time being sampled. The last community occupies only five percent of the list and hence should only occasionally be sampled. Four random numbers are drawn from a random number table between one and 200 to illustrate the selection process. In the example, the numbers are 73, 158, 166, and 195. If the number had been 73, it lies between 1 and 100 and identifies the first cluster as the one to be sampled. If the number was 158, the second cluster would be selected since 158 falls between 100 and 160. For a number of 166, the cluster would be selected, since the range on the cumulative list is 160 to 190. Finally if 195 was selected, the place on the cumulative list is between 190 and 200, or the spot occupied by the fourth cluster.

There are several ways to do *probability proportionate to size* sampling from a list, as observed in Figure 4-17. The first method is systematic PPS sampling. Here a *sampling interval* is derived by dividing the total population by the number of clusters to be sampled and then, after a random start in the first sampling interval, moving in a systematic manner

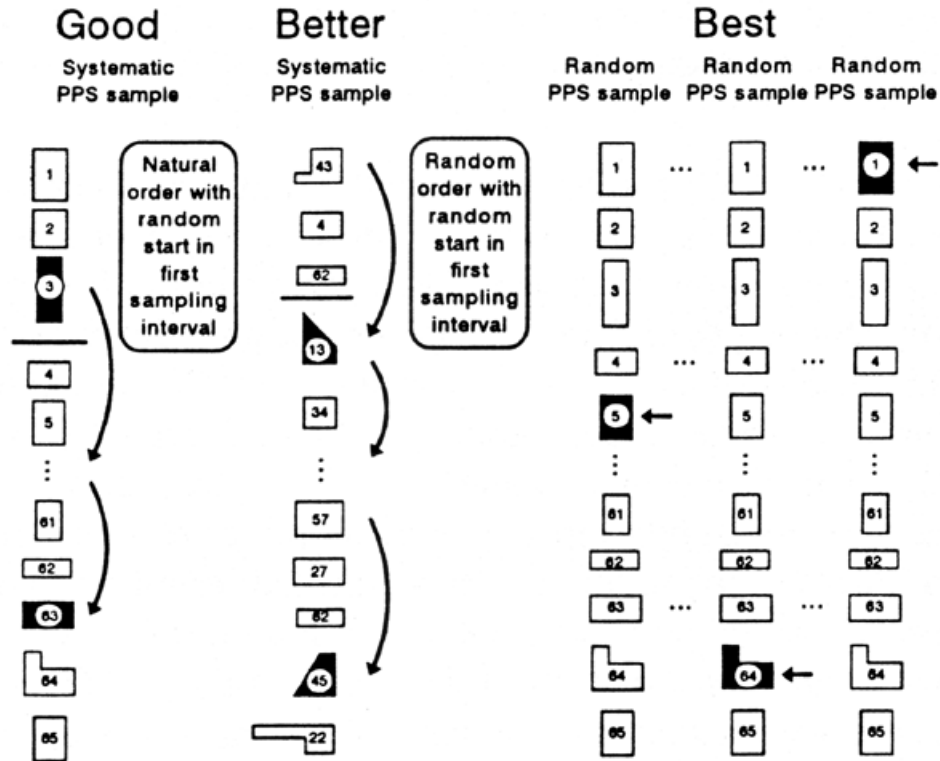


Figure 4-17. Methods of sampling with probability proportionate to size from a cumulative list of the population.

through the list of units.

The process is shown in Figure 4-18 for a population of 1,200 units from which 12 clusters are to be sampled. In Figure 4-18, part A, the population is listed in natural order in a cumulative list from one to 1,200. To derive the sampling interval, the population is divided by the number of clusters to be sampled, namely 12. This results in 12 smaller units of 100 units each (Figure 4-18, part B). There are several ways to draw a systematic sample from the population. One way is to start with the first unit in the first sampling interval and then move through the list in steps of the sampling interval (see Figure 4-18, part C). This results in 12 numbers,

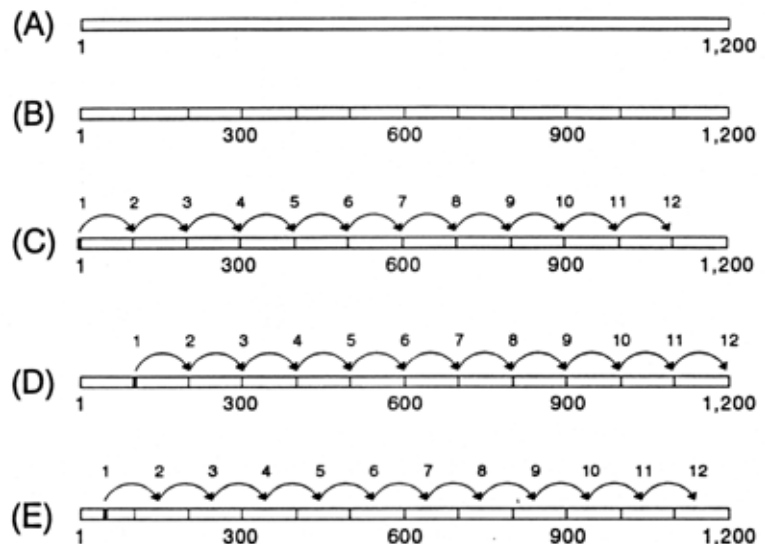


Figure 4-18. Steps for systematic PPS sampling from a cumulative list of the population.

starting with one and ending with 1,101. The communities where these units are located are the selected clusters. Unfortunately, with this method some units are always selected and others are never selected. For example, in the first sampling interval unit one is always selected and units 2 through 100 are never selected, resulting in a biased sample. Another approach would be to start with the last unit in the sampling interval and again move through the list in steps of the sampling interval (Figure 4-18, part D). This also is not very satisfactory since again, some units are always included while others are never included. Part E of Figure 4-18 shows the proper way to do this form of sampling. Here we select a random starting point in the first interval and then move in a systematic manner through the list, in steps of the sampling interval. Because of the random start, each unit has the same probability of being selected.

While *systematic PPS sampling* with a random start in the first sampling interval is an acceptable method of doing PPS sampling, it has two major disadvantages. The first is that all combinations of clusters cannot be selected. If two or more small clusters are listed in one sampling interval, there is no chance that both will be selected for inclusion in the study. The second disadvantage is that it is not clear if sampling is *with* or *without* replacement, and therefore whether we should or should not include the *finite population correction* term. While it may appear that this approach is sampling *without replacement*, clusters larger in size than the sampling interval will be selected two or more times, the same as sampling *with replacement*. Thus the sampling is actually a combination of *with* and *without* replacement.

A better approach is to first list in random order the various clusters before following the same systematic sampling scheme described above (see second column, Figure 4-17). Since the clusters are listed in random order, all combinations of clusters can be selected. Yet as before, large clusters will always be selected more than once if they exceed the size of the sampling interval. Because the scheme is a combination of sampling *with* and *without* replacement, we still have no firm guidelines on the use of the *finite population correction* term.

The best approach from a statistical point of view is draw a series of independent samples with *probability proportionate to size* from the cumulative list of clusters (see right section of Figure 4-17). The process is the same as described in Figures 4-15 and 4-16, using a random number table and a cumulative list of the units in each cluster. Earlier I described an example in Figure 4-18 of 12 clusters to be selected from a population of 1,200 units. To use the best selection method, we would draw 12 random numbers between one and 1,200 and select the corresponding clusters. Notice that all combinations of clusters can be selected, and that clusters may be selected more than once. Thus this approach is sampling *with replacement* and as a result, the *finite population correction* term would not be included in the variance equations.

4.4 ELEMENTARY UNITS

As mentioned earlier, the units to analyzed in a survey are termed *elementary units* or *elements*. Often in surveys, elementary units are also the units that are sampled, termed *sampling units*. An example of this is shown in Figure 4-19 for a survey of contraceptive practices among women. The *population* is all women in a geographic region of a country. The women are listed in a *frame*, and numbered in sequential order from one to the total number in the population. Women are sampled from the frame, and thus are the *sampling units*. Women are also analyzed as to their use of family

planning methods, and hence serve as *elementary units*.

In other surveys, the sampling unit is different from the elementary unit. Rather than women, an example of this second type of survey will feature men living in households (see Figure 4-20). Here the *population* to be sampled is households in a community. All households are listed in a *frame* and numbered consecutively. The *sampling unit* drawn from the frame is the household. Yet we are not interested in characteristics of the household. Instead we want to know more about the men who are in residence. In this survey, men are the *elementary units* to be analyzed, even though they were not sampled from a frame.

The distinction between sampling units and elementary units also occurs when doing two-stage surveys. For this third example we will view a two-stage cluster survey of families, shown in Figure 4-21. The *population* at the first stage is blocks in an urban area. The blocks serve as *clusters*. Starting at one, the blocks are listed in sequential order in a *frame* and a sample of blocks is selected.

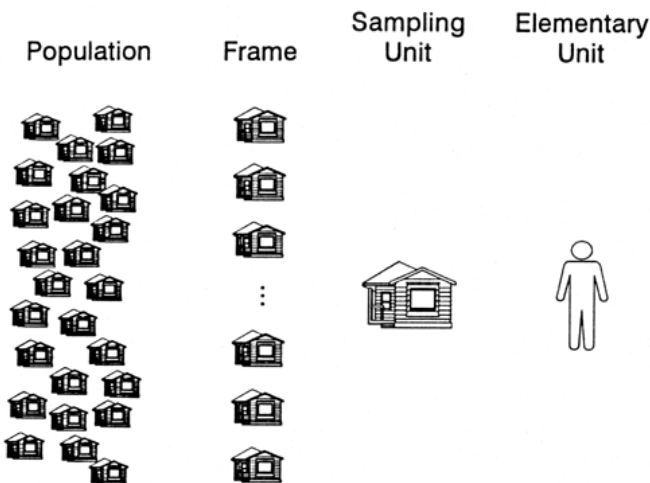


Figure 4-20. Sample of households from a population of households with a person serving as the elementary unit.

Chapter 2.

If a sample is drawn in an unbiased manner so that each person has the same probability of being selected, then the sampling unit is representative of a fixed number of units in the population. For example, if we sample 20 women from population of 10,000 women, each woman in the sample

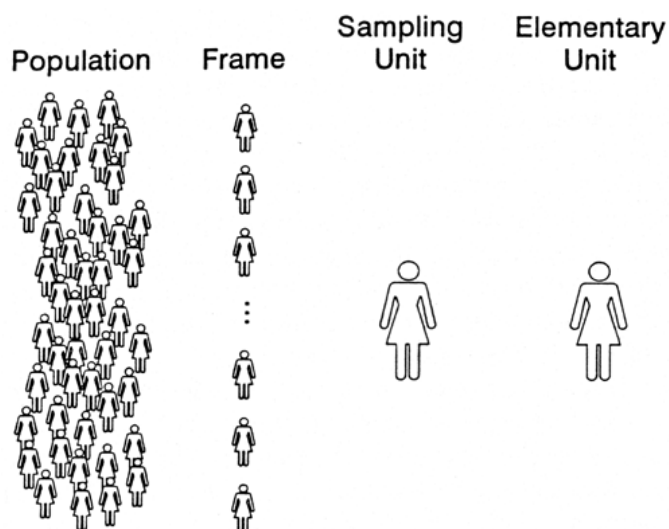


Figure 4-19. Sample of persons from a population of people with a person serving as the elementary unit.

This completes the first stage of the sampling process. In the blocks that are sampled, all households are listed and included in a second *frame*, but this of households rather than blocks. A *sample* of households is drawn from the list. Thus the selected households in the selected blocks are the *sampling units* of this survey. Again, however, we are not interested in households per se. Instead we want information on people living in the households. Therefore in this survey we interview or examine all persons who current live in the selected households, regardless of number. These people are the *elements* of the survey, similar in concept to the drug injections of addicts, featured in

on average represents 500 women in the population. The same holds true for households, as illustrated in Figure 4-22. Each household included in a sample is representative *on average* of a fixed number of households in the population. This means that the people in each household are also representative *on average* of persons in a fixed number of households in the general population. Notice that I have emphasized the term *on average*. Since we are involved in a sampling process of few units from large populations with nearly an infinite number of possible samples, there is great variability at the individual unit level. Hence the statement holds only for the average experience of all possible samples, and not necessarily for the one sample we selected.

So what happens when we change the characteristics of the sampled households by restricting the number or type of persons to be analyzed as elementary units? There is opportunity for bias. For example, assume in our survey shown in Figure 4-22 that we have decided to only interview or examine one person per household. As a result we have changed the characteristics of the households so that their residents are no longer representative, *on average*, of the households in the general population (see Figure 4-23). The situation becomes more complicated because households with one person still represent single person households in the community, but households with more than one person do not. We will address this problem more fully in later chapters. For now, however, it is important to recognize that the problem exists, especially when the sampling unit is different from the elementary unit.

4.5 SUMMARY

Sampling is an important tool for learning something

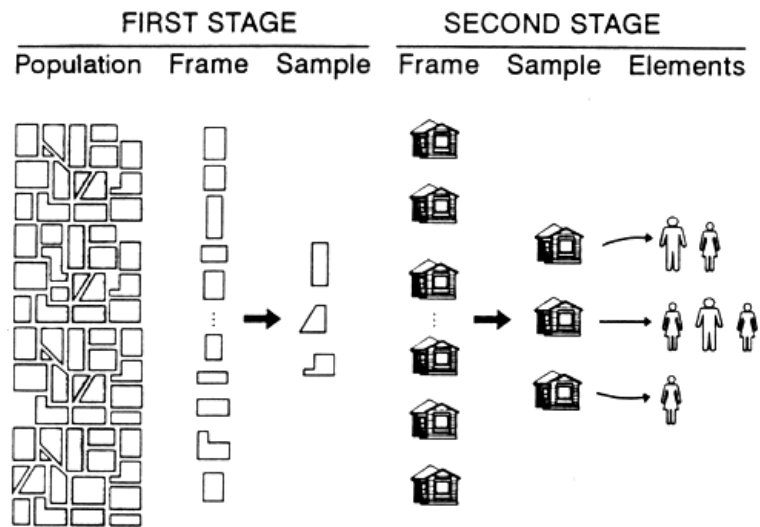


Figure 4-21. Two stage sampling of blocks at first stage and households at second stage with persons serving as elementary units.

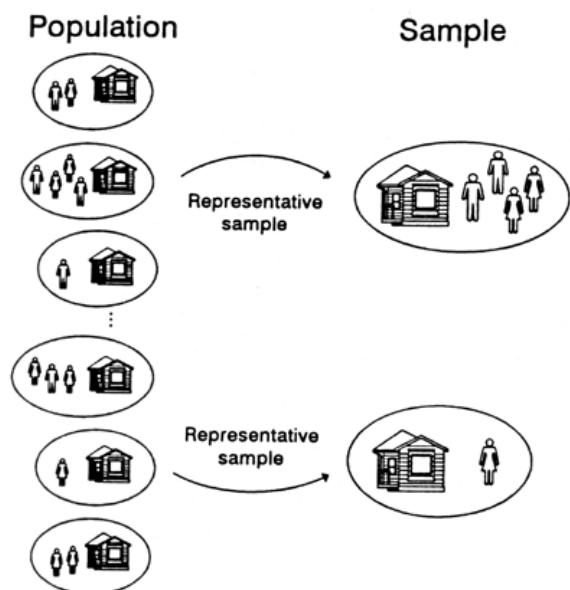


Figure 4-22. Households with residents in sample are representative of households with residents in population.

about a group without having to interview or examine all persons in the group. Yet if the sample is to be of value, each person in the population should have an opportunity to be included in the sample. When the opportunity is the same for all units in a population, we say that the sample is *self-weighted*. That is, each unit has the same probability of being selected, and all possible samples have the same probability of occurrence. While units in a population do not have the same chance of entering into a sample, we can use numeric *weights* to mathematically adjust the sample. These weights change the number of units in the population that the sampled units represent. Because weighting is more complicated, I prefer *self-weighted* samples for rapid surveys. There are, however, some situations where *weights* are useful, as will be described in Chapter 6.

In household surveys, elementary units (i.e., people) are different from the sampling units (i.e., households). If restrictions are placed on the elementary units in terms of number per household or some other criteria, then the sample may no longer be representative of the population. When this occurs, each person in the population no longer has the same probability of selection, and the sample may be biased. Ways to prevent this bias will be presented in future chapters.

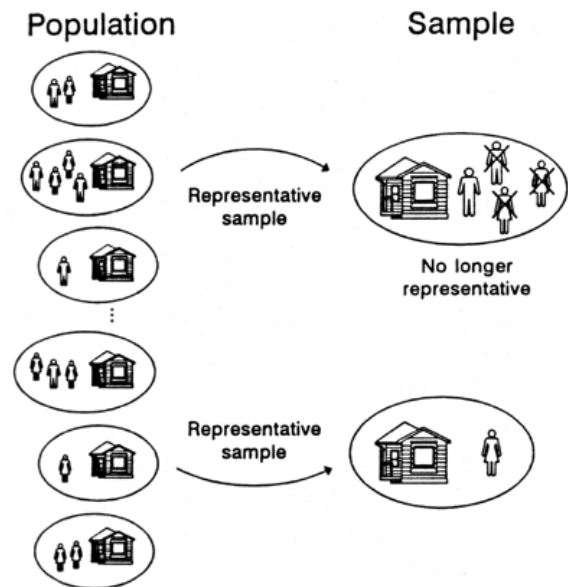


Figure 4-23. Households with residents in sample are not representative of households with residents in population.