Unit 15

Sampling Methods and Estimation of Sample Size

Contents

Learning Objectives

It is expected that after reading Unit 12 you would be able to

- ❖ Define what is sampling
- ❖ Classify sampling methods
- ❖ Calculate sample size.

15.1 Introduction

Unit 15 deals with the procedure of sampling◎ that helps you arrive at a subset of the universe of your research. It discusses the various methods of sampling and tells you how to work out a sample size. You will again read about sampling in Block 6. This is a subject you will need to master carefully as no matter what type of research you wish to carry out, you will need to apply your skill of the craft of sampling.

15.2 Sampling

A sample is a subset of the population that represents the entire group. When the population (or universe) is too large for the researcher to survey all its members because of its cost, the number of personnel to be employed, or the time constraint, a small carefully chosen sample is extracted to represent the whole (see Figure 15.1). The sample, as drawn in Figure 15.1, is expected to reflect the characteristics of the population.

A well selected sample may provide superior results. For example, in a research where well-trained interviewers are required, it may be possible to get a few trained interviewers to collect a sample rather than to get many trained interviewers to investigate the entire population. The trained interviewers may gather better quality information than non-trained or less trained interviewers. By contrast, if the population is sufficiently small, the entire population should be studied. When data are gathered on each and every member of the population, the study is known as a census study. The researcher is expected to clearly define the target population.

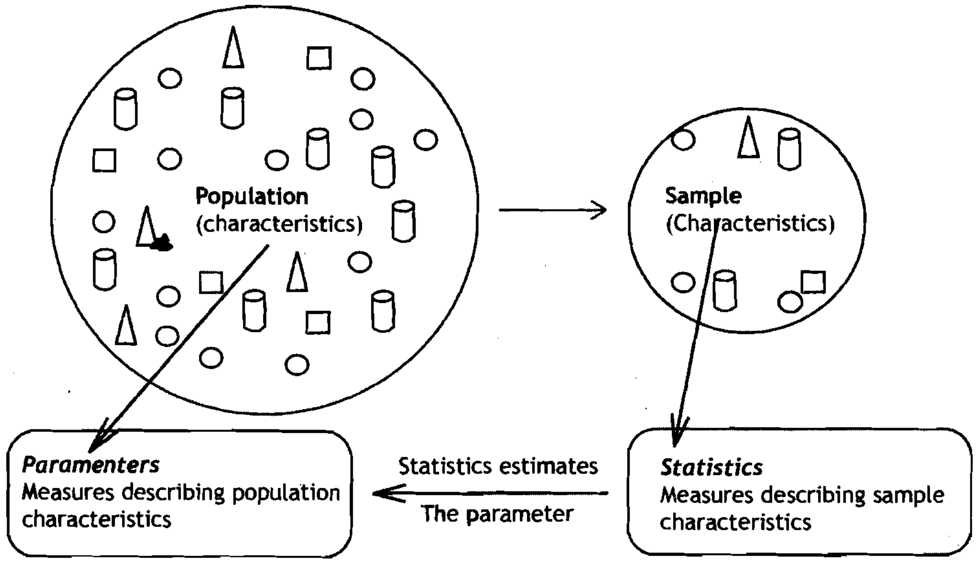


Figure 15.1

Relationship between Population, Parameter, Sample and Statistics

A population may be defined as an aggregate of individuals possessing a common trait or traits. There are two important factors: first that a population is the complete group about which knowledge is sought, and second each and every individual has some certain specified attribute or attributes.

Let us now complete Reflection and Action 15.1.

> **Reflection and Action 15.1**
>
> Work out the relationship between population, parameter, sample and statistics to reflect the characteristics of the population of the unit of your research project.

15.3 Classification of Sampling Methods

Sampling methods are classified into Probability or Non-probability. If the purpose of research is to draw conclusions or make predictions affecting the population as a whole (as most research usually is), then one must use probability sampling. But, if one is only interested in exploring how a small group, perhaps even a representative group, is doing for purposes of illustration or explanation, then one may use non-probability sampling.

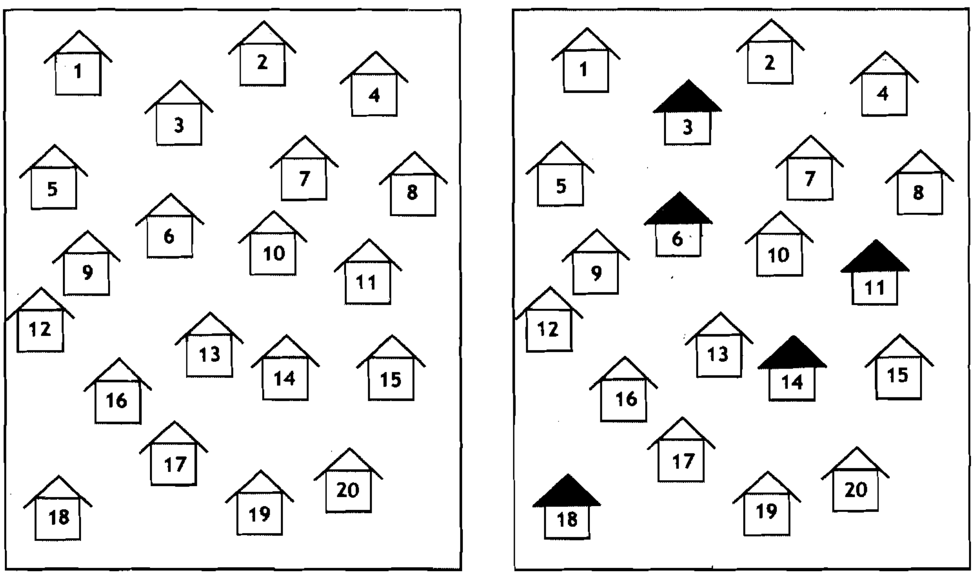
Let us first discuss probability sampling.

(A) Probability Sampling

In probability samples, each member of the population has a known non-zero probability of being selected. The key point behind all probabilistic sampling approaches is random selection. The advantage of probability sampling is that sampling error can be calculated, which is the degree to which a sample might differ from the population. Probability methods include random sampling, systematic sampling, and stratified sampling.

We shall discuss each of them.

a) **Random sampling** is the purest form of probability sampling. Each member of the population has an equal and known chance of being selected. The prerequisite for a random sample is that each and every item of the universe has to be identified. Random selection is effective in a clearly defined population that is relatively small and self-contained. When the population is large, it is often difficult or impossible to identify its each and every member, so the assemblage of available subjects becomes biased. One obtains a list of all residents or the voters list or telephone directory, and then selects a sample using a sequence of numbers from a random numbers table. Random numbers can also be created in numerous computer softwares. See Figure 15.2 that illustrates the selection of sample using random number table.



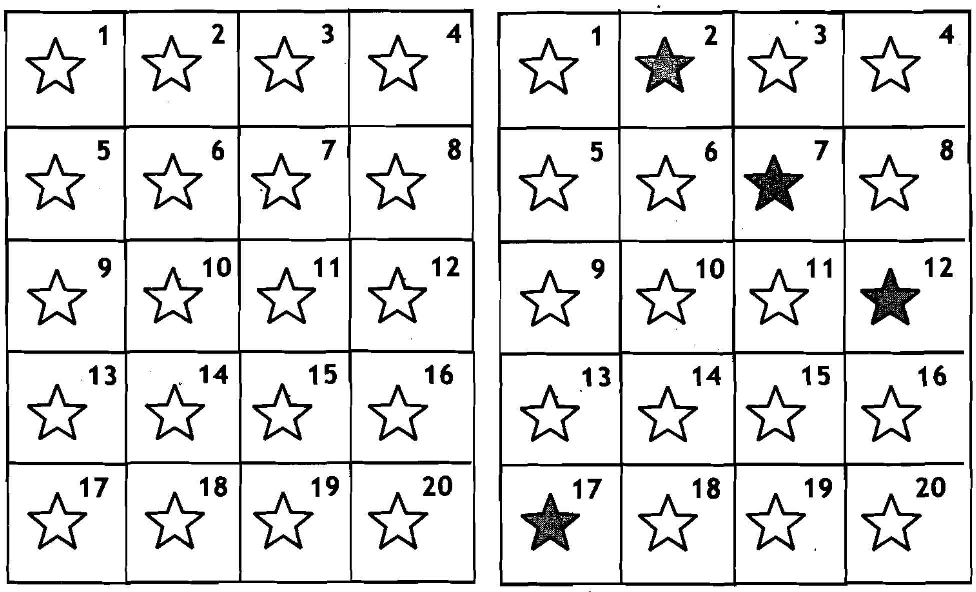
Population

Simple Random Sample drawn
(dark colour)
using Random number table

**Figure 15.2 Population Simple Random Sample drawn
(dark colour) using Random Number Table**

Source: Fisher, R. A. and F. Yates 1982. Statistical Tables. Longman: New York

b) **Systematic sampling** is also called an "Nth-name selection" technique. After the required sample size has been calculated, every Nth record is selected from a list of population members. As long as the list does not contain any hidden order, this sampling method is as good as the random sampling method. Its only advantage over the random sampling technique is simplicity. Systematic sampling is frequently used to select a specified number of records from a computer file. In Figure 15.3 you can find elucidation of the systematic random sampling method. The first number (2) has been selected by random number, followed by the selection of every 5th item in the series.

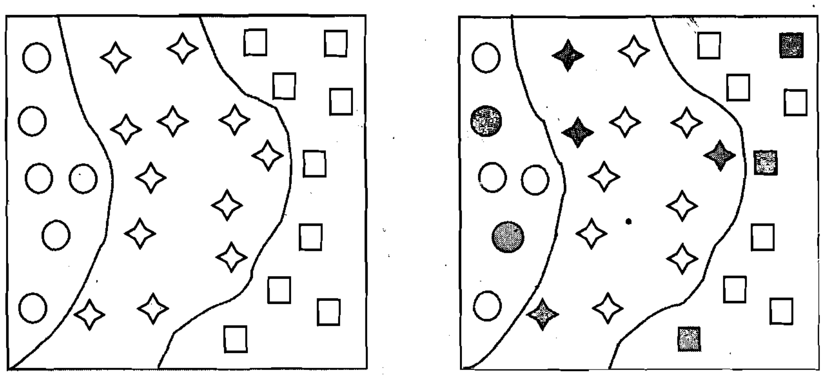


Population

Systematic Random Sample drawn (dark colour)

Figure 15.3 Systematic Random Sampling Method

c) **Stratified sampling** is a commonly used probability method that is superior to random sampling because it reduces the sampling error. A stratum is a subset of the population that shares at least one common characteristic. Examples of strata might be males and females, or managers and non-managers. The researcher first identifies the relevant strata and their actual representation in the population. Random sampling is then used to select a 'sufficient' number of subjects from each stratum. 'Sufficient' refers to a sample size large enough for the researcher to be reasonably confident that the stratum represents the population. Stratified sampling is most successful when (i) the within variance of each stratum is less than the overall variance of the population; (ii) when the strata in the population are of unequal size or have unequal incidence; and (iii) when sampling is cheaper in the strata. Figure 15.4 shows stratified random sampling method. Samples from the three strata have been extracted in proportion to their numbers.

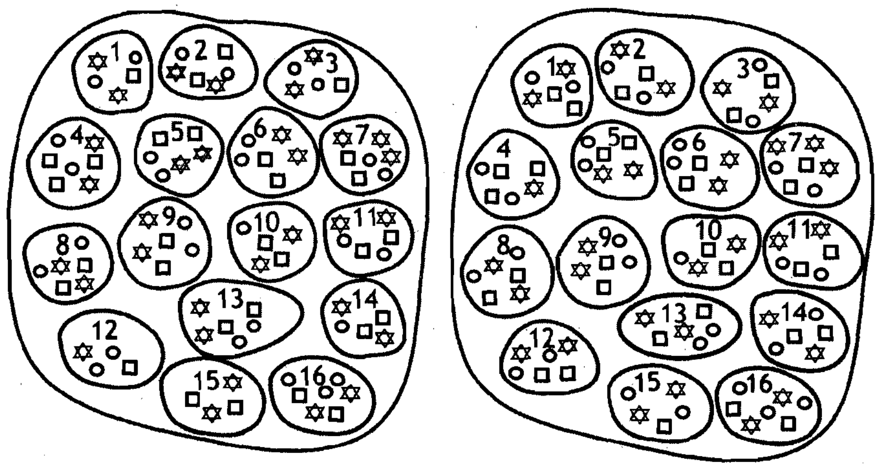


Population

Stratified Random Sample drawn (dark colour)

Figure 15.4 Stratified Random Sampling Method

d) **Cluster random sampling** is useful when the population is dispersed across a wide geographic region. This method allows one to divide the population into clusters and then select the clusters at random. Thereafter one can either study all the members of the selected clusters or again take random (simple or systematic) samples of these sampled clusters. If the latter system is followed, it is called multi-stage sampling. This method, for example, could be effective to study a tribal group or a community that is dispersed. The villages could be used as clusters and can be randomly selected. Figure 15.5 shows that five blocks (2, 7, 10 and 14) out of sixteen have been selected by random number. Each block contains a series of samples, as illustrated.



Population

Cluster Sample drawn (dark colour)

Figure 15.5 Cluster Random Sampling Method

> **Reflection and Action 15.2**
>
> Following the figures in the text, make figures based on the population pertaining to your research project that you selected while computing R & A 13.1 & 13.2 to show
>
> i) population simple random sample drawn in dark colour, using random number table
>
> ii) systematic random sampling method
>
> iii) stratified random sampling method
>
> iv) cluster random sampling method

(B) Non-probability Sampling

In non-probability sampling, members are selected from the population in some non-random manner. In this method, the degree to which the sample differs from the population remains unknown. Non-probability methods include Convenience sampling, Judgment sampling, Quota sampling and Snowball sampling. Let us now discuss each of the non-probability sampling methods.

a) **Convenience sampling** is used in exploratory research where the investigator is interested in getting an inexpensive approximation of the fact. As the name implies, the sample is selected because it is convenient. Also called haphazard or accidental, this method is based on using people who are a captive audience, just happen to be walking by, or show a special interest in research. The use of volunteers is an example of convenience sampling. This method is often used during preliminary research efforts to get a gross estimate of the results, without incurring the cost or time required to select a random sample.

b) **Judgment sampling** is a common non-probability method. The researcher selects the sample based on judgment. This is usually an extension of convenience sampling. For example, a researcher may decide to draw the entire sample from one 'representative' village, even though the population may be distributed over a number of villages. When using this method, the researcher 'feels' that the chosen sample is representative of the entire population.

c) **Purposive sampling**, much similar to judgment sampling, is where the researcher targets a group of people believed to be typical or average, or a group specially picked for some unique purpose. The researcher never knows if the sample is representative of the population, and this method is largely limited to exploratory research.

d) **Quota sampling** is the non-probability equivalent of stratified sampling. Like stratified sampling, the researcher first identifies the strata and their proportions in the population. Then convenience or judgment sampling is used to select the required number of subjects from each stratum. The researcher resorts to haphazard or accidental sampling, and makes no effort to contact people who are difficult to reach. This differs from stratified sampling, where the strata are filled by random sampling.

e) **Snowball sampling** is a special non-probability method used when the desired sample characteristic is rare. It may be extremely difficult or cost prohibitive to locate respondents in these situations. Snowball sampling relies on referrals from initial subjects to generate additional subjects. In other words, snowball sampling comprises identification of respondents who in turn refer researches to other respondents. This technique provides a means to access relatively invisible and vulnerable social groups. While this technique can dramatically lower the search costs, it comes at the expense of introducing bias because the technique itself reduces the likelihood that the sample will represent a good cross-section of the population. For example, an investigator finds a rare genetic trait in a person, and starts tracing his pedigree to understand the origin, inheritance and etiology of the disease.

You may have heard that only quantitative researches require sampling. The fact is that qualitative researches use sampling procedures (see Box 15.1).

Box 15.1 Use of Sampling in Qualitative Research

As Berger (1989) and Sarantkos have pointed out , it is fairly common for qualitative researches to use sampling procedures in the following manner.

i) Sampling is relatively small, dealing with typical cases.
ii) Use of flexible samples in size not requiring statistical calculations
iii) Use of purposive sampling dealing with non-probability
iv) Use of sampling to achieve suitability rather than representativeness
v) Sampling occurs while the research is in progress, rather than selecting a sample before starting it.

We would now focus on the procedure of calculating the sample size.

15.4 Sample Size

A prudent choice of the sample size for a particular survey involves many considerations, among which are the resources in manpower, cost per sample units and funds available, the number and type of parameters to be estimated. Obviously, these specifics will vary from one survey to another. All the same, a framework can be constructed within which general and viable decisions with respect to sample size can be taken. Sampling theory aids in arriving at good estimates of the sample size. The standard error here too provides the key.

Apart from the size of the universe the sample size may depend on the following conditions.

i) The confidence limit set up for estimation;
ii) The heterogeneity of the population; and
iii) Frequency/ proportion of the trait/ attribute to be examined.

The estimation of sample size also differs according to the purpose or the parameter under investigation. For example, whether sample size is being estimated for calculating mean, or proportion, or for comparing means. For illustration, let us consider, in Box 15.2 and Box 15.3 , two cases, namely, the estimation of the mean of a normally distributed variable and the estimation of a proportion. In these cases there are two assumptions, first that sampling is simple, random and without replacement, and second, the population sampled is infinitely large.

Box 15.2 Case One

<u>The sample size when estimating mean</u>

It is known that the standard error of mean can be calculated from the following formula.

SE_x = ó / vn 1

Where SE_x is the standard error of mean, ó is the standard deviation, and n is the sample size. Thus one can calculate sample size (n) using the following equation derived from equation 1.

n = (ó / SE_x)^2 2

Sample size can be calculated using the following steps.

Step 1: One requires the standard deviation of the universe, which is unknown. A rough estimate of this measure, however, is sufficient for suggesting sample size.

a) In many instances, the experience with similar problems will be a good guide for making this estimate of the standard deviation.

b) In other instances, an exploratory sample study on a small scale may be conducted in order to arrive at an estimate of ó.

c) To estimate the standard deviation of the universe, the range of the values in the universe may be estimated and used as a guide. It is known that in normal distribution the range is about six times the standard deviation. For practical purposes, an estimate of somewhere around one-fifth of the estimated range is often used.

   Suppose the range is roughly 300; that is, the difference between the lowest value in the universe and its highest value is 300. One-fifth of this rough estimate is 60. Therefore, one may take 60 as a rough approximation of ó.

Step 2: It must be decided how precise one wants the future sampling estimate to be. Thus, one may state that the estimate of the true mean is sufficiently precise if confidence limits of 12 are attached to it. Such an answer might be practicable for this particular problem.

Step 3: In this step the researcher has to decide the confidence limit. He may wish to be almost certain or be satisfied with, say, a 95% degree of confidence, that the specified limits will contain the true mean. The degree of confidence decided upon makes it possible to translate the interval decided upon in step 2 into standard error. If one is to be practically certain that true mean will lie within the interval of ±12 around the sample mean then the interval of ±12 becomes $\pm 3\ SE_x$. Therefore, $SE_x = 4$. If, on the other hand, one is willing to settle for a 95% degree of confidence, then ±12 becomes $\pm 2\ SE_x$ and $SE_x = 6$.

Using equation 2, the sample size for the above example will be

1) **Case 1-**
   At the level of practical certainty:
   Sample size (n) = $(60 / 4)^2 = 15^2 = 225$

2) **Case 2-**
   At the level of 95% confidence limit:
   Sample size (n) = $(60 / 6)^2 = 10^2 = 100$
   (In case 1 $SE_x = 4$, whereas in case 2 $SE_x = 6$)

Thus, in the above example, the sample size should be somewhere around 225 if one wishes to be practically certain that true mean will lie within an interval of ±12; but the sample need contain only 100 items if one settles for a 95% degree of confidence that true mean will lie within an interval of ±12.

Sometimes the acceptable difference between the sample and its true mean is expressed in percentage (say 3%) rather than absolute (as for example, ±12 in step 2 of the above example). Suppose the expected mean is around 500 then the acceptable interval would be ±15. But this necessitates an approximate knowledge of the expected mean.

Box 15.3 Case Two

Sample size when sampling for proportion

Consider the estimation of the proportion of individuals in a population with some particular attribute, for example those who own tractors for agriculture. This proportion, though not precisely known to the investigator, is generally known to him to an order of magnitude at least; that is to say, he will often know that owning tractors is quite rare (say, less than 3 in 1,000 persons), somewhat infrequent (3 in 100 to 3 in 1,000 persons), fairly common (3 in 10 to 3 in 100), or very common (more than 3 in 10). If owning tractors is known to be more infrequent than 3 in 100, a simple random sampling would invariably be much too inefficient and the other sampling methods appropriate to the estimation of rare events should be used. To assume random sampling amounts to assuming that the investigator's interest centers on only those attributes whose frequencies are at least 3 in 100. Even within these limits it is clear that if the population proportion is to be known exactly, the entire population must be examined. This is impracticable and generally unnecessary, for the investigator usually does not require this degree of exactness. His requirements are related, of course, to the use to which the estimate (or estimates) is to be put, and thus may vary from one investigator to another and with the proportion itself.

It is known that the standard error of a proportion can be calculated from the following formula.

$SE_p = v\ (PQ / n)$ 3

$SE_p^2 = PQ / n$ 4

Where, SE_p is the standard error of proportion, P is the proportion of an attribute in a population and Q is = 1 - P, and n is the sample size. Thus one can calculate sample size (n) using the following equation derived from equation 4.

$n = PQ / SE_p^2$ 5

Sample size can be calculated using the following steps:

Step 1: One requires an estimate of P, from which Q follows (Q = 1 - P), which is, of course, unknown. A rough estimate of this measure, however, is sufficient for suggesting sample size.

1) In many instances, experience with similar problems will be a good guide for making this estimate of the proportion.

2) In other instances, an exploratory sample study on a small scale may be conducted in order to arrive at an estimate of proportion.

If, however, neither of these two approaches is possible, then one can conservatively assume that P = 50% which leads to a larger sample size than any other value of P. This is because in a 50% - 50% break-up, the numerator (PQ) in the formula in equation 5 ($n = PQ / SE_p{}^2$), is the largest. However, for the following example, let us consider that P = 30% or 0.3.

Step 2: It must be decided upon how precise one wants the sampling estimate to be. The researcher may consider an interval of, say, ± 6% around a sample proportion as satisfactory in this situation.

Step 3: In this step the researcher has to decide the confidence limit. He may wish to be almost certain or be satisfied with, say, a 95% degree of confidence, that the specified limits will contain the true mean. In the former case, ± 6% will be equal to ± 3 SE_p and consequently SE_p =± 2%, whereas in the latter case, ± 6% will be equal to ± 2 SE_p and SE_p =± 3%.

Using equation 4, the sample size for the above example will be

1) Case 1 -
At the level of practical certainty:
Sample size (n) = $(0.3*0.7) / (0.02)^2 = 0.21 / 0.0004 = 525$

2) Case 2 -
At the level of 95% confidence limit:
Sample size (n) = $(0.3*0.7) / (0.03)^2 = 0.21 / 0.0009 = 233$
(P = 30% or 0.3; in case 1 $SE_p = .02$, whereas in case 2 $SE_p = .03$)

The use of a formula to obtain an estimate of sample size does not give us more than a rough approximation. In practice it is advisable to take the sample-size estimate as a bare minimum, to be increased for safety.

Let us now complete the Reflection and Action 15.3.

Reflection and Action 15.3

Suppose in your research project you wish to estimate sample size for calculating mean and the assumption is that sampling is simple and the population sampled is infinitely large. Further, you are in the stage of taking the three steps as elaborated in Case One given in the text, the exercise for you is to work out in detail each step and write it down in the fashion given just after Box 15.2.

15.5 Conclusion

Unit 15 discussed the important subject of sampling and provided you with relevant information on different methods of sampling. Further, it brought to you the skills of calculating the sample size.

You may like to keep in mind what Mitchell (1984: 239) said about sampling theory in statistics that it "devotes itself to providing numerical estimates of the likelihood that the population values be within some defined range of that established from the sample - provided that the sample has been chosen in such a way as to meet the mathematical conditions to justify the computation of the probabilities concerned." Further he clarified about another type of inference that is derived while using quantitative data to support theoretical interpretation and said, "The sophistication and elaboration for choosing a 'representative' sample in this restricted sense has overshadowed the other kind of inference involved when analytical statements are made from associations uncovered in a statistical sample. This is the inference that the theoretical relationship among conceptually defined elements in the sample will also apply in the parent population. The basis of an inference of this sort is the cogency of the theoretical argument linking the elements in an intelligible way rather than the statistical representativeness of the sample."

Further Reading

Burgess, R.G. (ed) 1982. *Field Research: A Sourcebook and Field Manual.* (Contemporary Social Research 4). George Allen and Unwin: London (Read page 76 onward fro discussions of random and non-random sampling)

Denizen, N.K. (ed.) 1970. *Sociological Methods: A Sourcebook.* Butterworths: London (Read page 81 onward for useful information on sampling techniques).