Probability and Non-probability sampling

There are certain issues to be taken into consideration while deciding to use probability or non-probability samples. Some research studies are not designed to be generalized to the population but collect exploratory data for designing questionnaires or measurement instruments. A non-probability sample is appropriate in these situations. Secondly, if the cost of a probability sample is too high in relation to the type and quality of information collected, then a non-probability sample is a possible alternative. Since, probability samples are often time consuming, a non-probability sample may be adopted to meet the time-constraints.

Although non-probability may be appropriate in certain situations, it is always best to use a probability sample when a study is conducted to support or refute a significant research question or hypothesis and the results will be generalized to the population.

We take up these two designs separately:

Non-probability sampling: In this type of sampling the items for the sample are selected deliberately by the researcher. In other words, the researcher purposively chooses particular units of the universe for constituting a sample. Mass media researchers frequently use non-probability sampling, particularly in the form of available samples, samples using volunteer subjects and purposive samples. Some of the different types of non-probability samples are:

   a) Accidental samples
   b) Available/Convenience samples
   c) Volunteer samples
   d) Purposive samples
   e) Quota samples

Accidental samples: In accidental sampling, the researcher simply reaches out and selects the subjects that he comes across and continues doing so till such time as the sample reaches a designated size. For example, he may take the first 150 persons he meets at a mall entry point who are willing to be interviewed or to provide the information he is seeking. In such a sample, there is no way of estimating bias except by doing a parallel study with a probability sample or undertaking a complete census. This does not mean that accidental samples have no place in scientific research. Besides being economical and convenient, they can provide a basis for stimulating insights and hypotheses.

Available samples: An available sample is also known as a convenience sample. It is a collection of readily accessible subjects for study, such as a group of students enrolled in a mass media course or shopkeepers in a mall. Although available samples are helpful in collecting exploratory information, the samples may contain unknown quantities of error. Researchers need to consider both the positive and negative aspects of available samples before using them in a research study. Available samples are a subject of debate in research. Critics argue that available samples do not represent the population and therefore have no external validity. Proponents of available samples claim that if a particular trait or
characteristic does exist, then it should exist in any sample. Available samples can be useful in pretesting questionnaires or conducting a pilot study.

**Volunteer samples**: Persons who willingly participate in research projects are known as volunteer samples. Subjects who constitute a volunteer sample also form a non probability sample as the individuals are not selected according to mathematical guidelines. Researchers have found that volunteer subjects tend to exhibit higher educational levels, occupational status and intelligence levels. These characteristics imply that the use of volunteer subjects may significantly bias the results of the research and may lead to inaccurate assumptions of various population parameters. In some cases volunteer subjects are necessary but they should be used carefully since they contain unknown quantity of error. Volunteer samples are extensively used these days by the media and internet websites. Various polls conducted on radio and television stations, TV networks, the Internet, newspapers and magazines use volunteer samples. However, volunteer samples are shown to be inappropriate in scientific research.

**Purposive samples**: The basic assumption behind purposive sampling is that the subjects are selected for a specific characteristic or quality and eliminates those who fail to meet these criteria. Purposive samples are often used in advertising studies where researchers select subjects who use a particular type of product and ask them to compare with a new product. However, in such a sampling there is no assurance that every element or subject has some specifiable chance of being selected. Here, the sampling errors and biases cannot be computed since the sampling procedure does not involve probability sampling at any stage.

**Quota samples**: One of the most commonly used methods of sampling in market research is the method of quota sampling. Here the subjects are selected to meet a predetermined or known percentage. The basic objective of quota sampling is the selection of a sample that is similar to the population in terms of proportion of certain characteristics. For example, a researcher is interested in finding out how girl students differ from boys in their intelligence levels in a co-educational institution. And, there is a sharp difference in the proportion of girls and boys studying in the institution, then, a quota sample is appropriate in order to reflect the population characteristics. In quota sampling the population is reflected in terms of certain characteristics and the proportion of the population with specific characteristics is determined and selected like-wise.

**Probability Sampling**: This type of sampling corporate a systematic selection procedure to ensure that each unit has an equal chance of being selected. However, it does not always guarantee a representative sample from the population, even when systematic selection is followed. It is possible to randomly select 50 students of a university hostel in order to determine the average number of hours spent on watching television during a typical week and discover that there was no TV set installed in the hostel or even if it was installed it was never in a working condition. This may be unlikely but it underscores the possibility to replicate any study.

The most commonly used probability samples are:
a) Simple random samples  
b) Systematic random samples  
c) Stratified random samples  
d) Cluster samples

**Simple random sampling:** The most basic type of probability sampling is the simple random sampling. Here, each subject or unit in the population has an equal chance of being selected. In principle, one can use this method from selecting random samples from populations of any size. But in practice, it becomes very cumbersome.

If a subject or unit is drawn from a population and removed from subsequent selections, the procedure is known as random sampling *without replacement* - a widely used random sampling method. Random sampling *with replacement* involves returning the subject or unit to the population so that it has an equal chance of being selected another time.

**Table of random numbers:** Researchers also use the list of random numbers to generate a simple random sample. For example, a researcher wants to analyse the portrayal of women in 10 soap operas on television channels out of a population of 100 programs then he can use the table of random numbers (Table 9.2) to select 10 programs by numbering each of the 100 programs from 00 to 99. First a starting point in the table is selected. There is no specific way to choose a starting point; it is the discretion of the researcher. The researcher then selects the remaining 9 numbers by going left, right, up or down. For example, if the researcher goes down the table from the starting point 39 then his drawn sample will include programs numbered 39, 02, 78, 94, 71, 83, 20, 49, 64, 08 and 55.

Simple random samples for use in television surveys are often obtained by a process called random digit dialling. This method involves the randomly selected four-digit numbers and adding them to the three-digit or four-digit exchange prefixes in the city in which the survey is conducted. Many of the telephone numbers generated by this method are invalid because some phone numbers are disconnected or they may be temporarily out of service and so on. Therefore it is best to consider three times the number of telephone numbers needed; if a sample of 100 is required then at least 300 telephone numbers should be generated.

<table>
<thead>
<tr>
<th>Table of Random numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 33 04 81 00 95 62 79 94 07 12 85</td>
</tr>
<tr>
<td>09 50 23 08 48 37 49 96 10 11 03 14</td>
</tr>
<tr>
<td>10 19 16 47 37 21 44 52 02 55 18 77</td>
</tr>
<tr>
<td>04 54 22 12 39 43 57 79 83 86 05 13</td>
</tr>
<tr>
<td>99 00 60 35 28 95 80 20 66 00 02 59</td>
</tr>
<tr>
<td>55 94 58 98 83 58 68 31 49 79 73 15</td>
</tr>
<tr>
<td>49 96 10 11 03 14 73 88 39 03 19 29</td>
</tr>
</tbody>
</table>
Random number generation is possible through a variety of methods. However, two basic rules must be kept in mind: (1) each subject in the population must have an equal chance of being selected (2) The selection process must be free from bias of the researcher.

The purpose of random sampling is to reduce sampling error and overlooking the above mentioned rules only increases the chance of error creeping into the study.

Simple random samples for use in television surveys are often obtained by a process called random digit dialing. This method involves the randomly selected four-digit numbers and adding them to the three-digit or four-digit exchange prefixes in the city in which the survey is conducted. Many of the telephone numbers generated by this method are invalid because some phone numbers are disconnected or they may be temporarily out of service and so on. Therefore it is best to consider three times the number of telephone numbers needed; if a sample of 100 is required then at least 300 telephone numbers should be generated.

Random number generation is possible through a variety of methods. However, two basic rules must be kept in mind: (1) each subject in the population must have an equal chance of being selected (2) The selection process must be free from bias of the researcher.

The purpose of random sampling is to reduce sampling error and overlooking the above mentioned rules only increases the chance of error creeping into the study.

**Systematic Random sampling:** The most practical way of sampling is to select every i\text{th} item on a list. Sampling of this type is known as systematic random sampling. For example, to obtain a sample of 50 from a population of 500, or a sampling rate of 1/10, a researcher
randomly selects a starting point and sampling interval. Thus, if the number 4 is chosen as the starting point, then the sample will include the 50 subjects or items numbered 4, 14, 24, 34, 44, 54, and so on. To add further randomness to the process, the researcher can select the starting point and the sampling interval randomly. For example, an interval of 6 with a starting point 37 generates the numbers 43, 49, 55, 61, and so on.

In systematic random sampling selection of subjects is easy, accurate and generally inexpensive as compared to simple random sampling. However, a complete list of the population must be obtained and hidden periodicity in the population may bias the process.

**Stratified random sampling:** If a population from which a sample is to be drawn does not constitute a homogeneous group, stratified sampling design is generally used in order to obtain a representative sample. In some cases, the researcher wants to ensure that a specific sub sample gets adequate representation. In such instances, a stratified random sampling approach is adopted to get an adequate representation of the sub sample. The population is divided into different strata with specific characteristics. The characteristics or features of the strata or segment may include any variable: age, gender, religion, occupation, income level etc. There are some benefits of stratification: (1) Sampling from strata ensures that at least some elements of the total sample will come from each strata (2) Since the procedure involves choosing independent samples from each strata, different methods of selection can be used for one or more strata (3) Sampling error is considerably reduced because each strata has an assured representation in the sample.

Stratified random sampling can be applied in two different ways. *Proportionate stratified sampling* includes strata with sizes based on their proportion in the population. If 20% of the population is in their teens in the age-group of 13-19, then 20% of the total sample will be subjects in this age-group. This gives each person in the population an equal chance of being selected. In *disproportionate stratified sampling*, the sample units are selected on equal basis from each stratum irrespective of their size of representation of their strata in the population.

Some merits of the stratified random sampling method are that selection of the sample units is made from a homogeneous group and comparisons can be made to other populations. One of the disadvantages of this method is that this method is time-consuming as well as expensive.

**Cluster Sampling:** A key difference between cluster sampling and stratified random sampling is that the total population is divided into clusters or groups or categories and the sample units are randomly selected from each of these clusters. The ultimate sample consists of all units from these clusters. With cluster sampling, the researcher can divide the state into districts or zip code areas and select groups from these areas. For example, a zip code area may contain mostly residents of a high socio-economic status who are unrepresentative of the rest of the state; if selected for analysis, such a group may confound the results. To help control such error, it is best to use small areas or clusters, both to decrease the number of elements in each cluster and to maximise the number of clusters selected (Babbie, 2001).If
clusters are formed or based on some geographic sub-divisions, then cluster sampling is better known as area sampling.

**Multi-stage sampling:** Multi-stage sampling is a further development of the principle of cluster sampling. In many nation-wide surveys, researchers use a form of cluster sampling called multi-stage sampling, in which individual households or persons and not groups are selected. In multi-stage sampling, each stage is a separate sampling task. Therefore, the separates stages can be approached with different methods. For example, in a large population, stratification may be useful within the initial steps and simple random procedures in final ones.

**Sample Size:** Every sampling procedure has to deal with the question of how big or large a sample to draw. Unfortunately this simple question does not have a simple answer. The size of the population is dependent on certain considerations of the project under study. (i) The resources and time available (ii) the complexity of the characteristics of the population under study (iii) the precision required to approach the study.

The complexity of a characteristic depends on the number of categories used to measure it. Say for example, the likelihood of a respondent watching a reality show on television. One can begin asking “Yes/No” question. Then we may investigate the “Yes/No” responses by the sex of the respondent, 3 levels of socio-economic class and 6 levels of education and so on. In this case we can have a 2x2x3x6 category scheme or 72 different cells.

Precision is determined by the efficiency of the statistical procedure to be used and the variability of the characteristic under study. It does not involve the size of the population. Here, the researchers’s time and money are important.

Characteristics which are complex and must be approached with high precision require large samples and considerable resources. A rough rule of thumb for the novice researcher is that samples of less than 30 are generally considered inadequate for pre-testing; samples in the 100 to 200 range are rarely brought into question on the basis of size and some research questions require samples of greater than 500.

**Sampling Bias:**

Sampling bias refers to a sampling procedure which systematically excludes some member or members of the population. Whereas convenience and judgment samples are by definition biased samples, true random samples are freed from sampling bias but not sampling error by their nature. Most probability samples are not truly random, however, because of the difficulty of effecting random selection. Sampling bias is also the consequence of respondents refusing to participate in a study even when drawn in a sample.

**References and Further Readings**


