4. Sampling

4.1 Chapter summary

In this chapter we will focus on a very important construct in the field of marketing research, sampling. The chapter will start with a discussion on the importance of sampling in marketing research which will be followed by understanding some basic constructs and terms used by researchers in the field of sampling. The chapter will also discuss briefly on how to determine the sample size. Both probability and nonprobability methods will be discussed in details in this chapter with advantages and disadvantages associated with each technique. It will also focus on what criteria should be kept in mind when selecting an appropriate sampling technique.

4.2 Importance of sampling in marketing research

Sampling is one of the very important aspects of marketing research. From a general perspective, sampling involves selecting a relatively small number of elements (characteristics) from a larger defined group of elements and expecting that the information gathered from the small group of elements will provide accurate judgement about the larger group. We use sampling in our decision making almost every time. For example, before buying a book we flick through few pages and decide weather it suits our reading preferences. For a complex buy such as a mobile phone, we first decide several features as essential and others as desirable. Then we decide on the brand and select the mobile phone on the brand, price of the product and several other such variables. While making the final decision there are many such variables which we don't take into consideration. In a way, we use few elements (characteristics) of mobile phone (or a book) and expect that they will cover most of what we desire. We use sampling when selecting a job, choosing a restaurant and even selecting TV channels. As we consumers use sampling in our regular decision making, managers can also benefit by understanding sampling process in providing better matched products with our needs.

Almost every newspaper everyday reports the results of studies in which public opinion on some question is estimated by collecting opinions from a few selected individuals. Much marketing information is obtained in a similar fashion, using a sample of consumers. Therefore, it is very important for a market researcher to understand the concept of sampling. Furthermore, sampling provides several benefits overall. For example, as not every consumer of the product is being studied, the total cost of research can be lowered with the use of sampling. A sample would require fewer fieldworkers. Therefore, better personnel could be selected and trained and their work could be closely supervised. It is observed that the lesser administrative problems encountered in collecting data from a sample lead to more accurate data than could be obtained by collecting data from all units.[43]

4.3 Sampling: basic constructs

As we defined sampling above, there are several other constructs which need defining before delving deeply into the phenomenon of sampling. Sampling is conducted when conducting a census is impossible or unreasonable. The studies which cover all the members of population are called 'census' which are generally carried out by national governments in various countries. Most countries carry out such surveys every 10 years. Census studies involve the population overall. In research terms, 'population' is defined as the totality of cases that confirm to some designated specifications.[44] For example, if a manager of brand X of washing machine was interested in understanding customer satisfaction relating to washing machines, the researcher will need to study all consumers who owned a washing machine (i.e. population) to get an accurate idea. However, studying population will be unreasonable in this case because the number of people owning washing machine will be huge and so the study will require unreasonable amount of resources in terms of cost and time. Most managers that require research data for decision making are not interested in total population response, but rather with a prescribed segment of the total. Such prescribed segments are defined as 'target population'. A target population consists of the complete group of elements (people or objects) that are specifically identified for investigation according to the objectives of the research project.[45] Continuing the earlier example, the defined target population for the washing machine study will be washing machine owners of brand X.

A precise definition of the target population is essential and usually done in terms of 'elements', 'sampling units' and 'sampling frame'. An element is defined as a person or object from which data is sought and about which inferences are to be made. For example, target population elements for the washing machine study might include a particular brand (i.e. Brand X); specific group of people (i.e. females). Sampling units are the target population elements available for selection during the sampling process. Using the washing machine example, a sampling unit may be females who have purchased new washing machines rather than a second hand one. Choice of elements and sampling units may redefine the study. In case of washing machine it may now change from 'customer satisfaction among washing machine owners' to 'customer satisfaction among new brand X washing machine owner females'. The above example gives a brief overview of selecting target population, elements and sampling unit. However, in real life, deciding a target population is a highly complex task[46] as many other variables are involved.

A sampling frame is a representation of the elements of the target population. It consists of a list or set of directions for identifying the target population. Some common sources of sampling frame are lists of voters, commercial directories, telephone directories, or even maps. Many commercial organizations provide a database consisting of names, addresses, and telephone numbers of potential sampling frame for various studies. Regardless of the sources, it is very difficult and expensive to obtain truly accurate or representative sampling frames.[47] For example, it will not be easy to obtain the addresses and names of new washing machine owners. However, in comparison it will be very difficult if the study was focused on second hand washing machine owners.

Such difficulties in obtaining an accurate sampling frame leads to 'sampling frame error'. It can be defined as the variation between the population defined by the researcher and the population used. For example, telephone directories can be a source for such errors as it does not provide unlisted numbers or numbers which are obtained after the publication dates. At the same time it does provide numbers which might be cancelled or disconnected.

Throughout the research process a researcher can make errors in judgement that results in creating some type of bias. All such types of errors are classified in marketing research as sampling or nonsampling errors. Sampling errors represent any type of bias that is attributable to mistakes in either drawing a sample or demining the sample size. This leads to the sample being non-representative to the population and is at times called random sampling error also. Nonsampling errors represent a bias that occurs regardless of sample or census being used. Nonsampling errors can be categories as nonresponse error (respondent is unable or unwilling to respond) or response errors (inaccurate, misreported or misanalysed response).

4.4 Determining sample size

Determining sample size is a complex task and involves much clarity with regard to the balance between the resources available and number or accuracy or information obtained. Since data collection is generally one of the most expansive components of any research project various factors play a crucial role in determining the final sample size. Several qualitative and quantitative factors are considered when determining the sample size. The qualitative issues considered may include factors such as: (a) nature of research and expected outcome; (b) importance of the decision to organization; (c) number of variables being studied; (d) sample size in similar studies; (e) nature of analysis and (f) resource constraints. Various quantitative measures are also considered when determining sample size such as: (a) variability of the population characteristics (greater the variability, larger the sample required); (b) level of confidence desired (higher the confidence desired, larger the sample required); and (c) degree of precision desired in estimating population characteristics (more precise the study, larger the sample required).

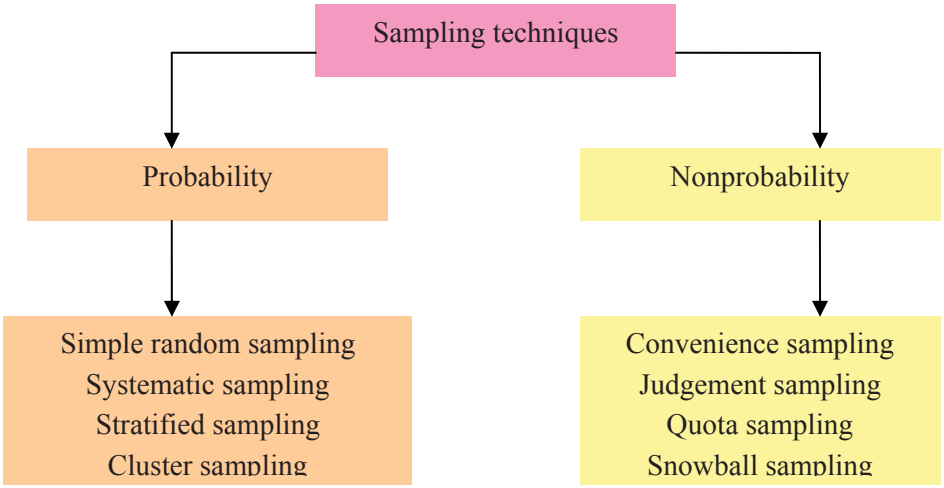
The size of sample also depends on the type of study that is being undertaken. Problem identification research (as defined in chapter 1) may require a sample of 1000 in comparison to problem solving research in the range of 300-500.

4.5 Classification of sampling techniques

How to obtain a sample is an important issue relating to research design. There are two basic sampling designs: probability and nonprobability sampling design. Of these two techniques, probability sampling is more robust in comparison as in this technique each sampling unit has a known, nonzero chance of getting selected in the final sample. Nonprobability techniques on the other hand, do not use chance selection procedure. Rather, they rely on the personal judgement of the researcher. The results obtained by using probability sampling can be generalized to the target population within a specified margin of error through the use of statistical methods. Put simply, probability sampling allows researchers to judge the reliability and validity of the findings in comparison to the defined target population. In case of nonprobability sampling, the selection of each sampling unit is unknown and therefore, the potential error between the sample and target population cannot be computed. Thus, generalizability of findings generated through nonprobability sampling is limited. While probability sampling techniques are robust in comparison one of the major disadvantages of such techniques is the difficulty in obtaining a complete, current and accurate listing of target population elements.

Both probability and nonprobability sampling procedures can be further sub-divided into specific sampling techniques that are appropriate for different circumstances. Figure 4.1 provides details relating to the classification of sampling techniques.

Figure 4.1:
Classification of sampling techniques



In the following section we shall discuss each of the sampling techniques.

4.6 Probability sampling techniques

As stated in figure 4.1 probability sampling techniques can be classified into four sub-categories namely; simple random sampling; systematic sampling, stratified sampling and cluster sampling.

4.6.1 Simple random sampling

Simple random sampling is a probability sampling technique wherein each population element is assigned a number and the desired sample is determined by generating random numbers appropriate for the relevant sample size. In simple random sampling, researchers use a table of random numbers, random digit dialling or some other random selection methods that ensures that each sampling unit has a known, equal and nonzero chance of getting selected into the sample. For example, let us assume that the manager of the washing machine Brand X had the name and addressees of all new washing machine buying females (assume the total number is 1000). The manager could create a label associating with each person and put them in a big jar and select washing machine owners from the same. This way each washing machine owner female has an equal, nonzero chance of getting selected. If the number of owners was much larger a random number table can be used however, the chance of each owner getting selected still remains equal and nonzero.

4.6.2 Systematic random sampling

In systematic random sampling the sample is chosen by selecting a random starting point and then picking each *i*th element in succession from the sampling frame. The sampling interval *i*, is determined by dividing the population size *N* by the sample size *n* and rounding to the nearest integer. For example, if there were 10,000 owners of new washing machine and a sample of 100 is to be desired, the sampling interval *i* is 100. The researcher than selects a number between 1 and 100. If, for example, number 50 is chosen by the researcher, the sample will consists of elements 50, 100, 150, 200, 250 and so on.[48] In simple terms, systematic sampling is similar to the simple random sampling however requires that the target population be ordered in some way. Systematic random sample elements can be obtained via various means such as customer list, membership list, taxpayer roll and so on. This technique is frequently used as it is a relative easy way to draw sample while ensuring randomness. One of the drawbacks of this technique is that if a hidden pattern exists in the data the finding may not be truly representative of the target population. However, the potential small loss in overall representativeness is normally countered by significantly larger gains in time, effort and cost.

4.6.3 Stratified sampling

Stratified sampling is distinguished by the two-step procedure it involves. In the first step the population is divided into mutually exclusive and collectively exhaustive sub-populations, which are called strata. In the second step, a simple random sample of elements is chosen independently from each group or strata. This technique is used when there is considerable diversity among the population elements. The major aim of this technique is to reduce cost without lose in precision. There are two types of stratified random sampling; (a) proportionate stratified sampling and (b) disproportionate stratified sampling. In proportionate stratified sampling, the sample size from each stratum is dependent on that stratum's size relative to the defined target population. Therefore, the larger strata are sampled more heavily using this method as they make up a larger percentage of the target population. On the other hand, in disproportionate stratified sampling, the sample selected from each stratum is independent of that stratum's proportion of the total defined target population. There are several advantages of stratified sampling including the assurance of representativeness, comparison between strata and understanding of each stratum as well as its unique characteristics. One of the major difficulty however, is to identify the correct stratifying variable.

4.6.4 Cluster sampling

Cluster sampling is quite similar to stratified sampling wherein in the first step the population is also divided into mutually exclusive and collectively exhaustive sub-populations, which are called clusters. Then a random sample of clusters is selected, based on probability random sampling such as simple random sampling. The major difference between stratified and cluster sampling is that in stratified sampling, all the subpopulations (strata) are selected for further sampling whereas in cluster sampling only a sample of subpopulations (clusters) is chosen. The objectives of these methods are also different. The objective of stratified sampling is to increase precision while cluster sampling strives to increase sampling efficiency by decreasing costs. Because one chooses a sample of subgroups with cluster sampling, it is desirable that each subgroup be a small scale model of the population. Thus, the subgroups (clusters) ideally should be formed to be as heterogeneous as possible. If all elements in each selected cluster are included in the sample, the procedure is called one-stage clustering. However, if a sample of elements is drawn probabilistically from each selected cluster, the procedure is called two-stage clustering. The most common form of cluster sampling is area sampling in which the clusters consists of geographical areas. There are several advantages of cluster sampling including the reduction in costs due to available data with regard to population groups (such as telephone directories and address lists) and feasibility of implementation. However, one of the major disadvantages of cluster sampling is the homogeneity among the selected cluster. Ideally each cluster should represent the population at large however, in reality it is quite difficult to achieve.

4.7 Nonprobability sampling techniques

The selection of probability and nonprobability sampling is based on various considerations including, the nature of research, variability in population, statistical consideration, operational efficiency and sampling versus nonsampling errors. Nonprobability sampling is mainly used in product testing, name testing, advertising testing where researchers and managers want to have a rough idea of population reaction rather than a precise understanding. Ad depicted in figure 4.1 there are various types of nonprobability sampling including, convenience sampling, judgement sampling, quota sampling, snowball sampling.

4.7.1 Convenience sampling

As the name implies, in convenience sampling, the selection of the respondent sample is left entirely to the researcher. Many of the mall intercept studies (discussed in chapter 3 under survey methods) use convenience sampling. The researcher makes assumption that the target population is homogenous and the individuals interviewed are similar to the overall defined target population. This in itself leads to considerable sampling error as there is no way to judge the representativeness of the sample. Furthermore, the results generated are hard to generalize to a wider population. While it has a big disadvantages relating to sampling error, representativeness and generalizability, convenience sampling is least time consuming and least costly among all methods.

4.7.2 Judgement sampling

Judgement sampling, also known as purposive sampling is an extension to the convenience sampling. In this procedure, respondents are selected according to an experienced researcher's belief that they will meet the requirements of the study. This method also incorporates a great deal of sampling error since the researcher's judgement may be wrong however it tends to be used in industrial markets quite regularly when small well-defined populations are to be researched. For example, if a manager wishes to the satisfaction level among the key large-scale business customers judgement sampling will be highly appropriate. Same as convenience sampling, judgement sampling also has disadvantages relating to sampling error, representativeness of sample and generalizability however the costs and time involvement is considerably less.

4.7.3 Quota sampling

Quota sampling is a procedure that restricts the selection of the sample by controlling the number of respondents by one or more criterion. The restriction generally involves quotas regarding respondents' demographic characteristics (e.g. age, race, income), specific attitudes (e.g. satisfaction level, quality consciousness), or specific behaviours (e.g. frequency of purchase, usage patterns). These quotas are assigned in a way that there remains similarity between quotas and populations with respect to the characteristics of interest. Quota sampling

is also viewed as a two-stage restricted judgement sampling. In the first stage restricted categories are built as discussed above and in the second stage respondents are selected on the basis of convenience of judgement of the researcher. For example, if the researcher knows that 20% of the population is represented by the age group 18-25, then in the final sample s/he will try to make sure that of the total sample 20% of them represent the age group 18-25. This procedure is used quite frequently in marketing research as it is easier to manage in comparison to stratified random or cluster sampling. Quota sampling is often called as the most refined form of nonprobability sampling.[49] It also reduces or eliminates selection bias on the part of field workers which is strongly present in convenience sampling. However, being a nonprobability method it has disadvantages in terms of representativeness and generalizability of findings to a larger population.

4.7.4 Snowball sampling

In snowball sampling, an initial group of respondents is selected, usually at random. After being interviewed however, these respondents are asked to identify others who belong to the target population of interest. Subsequent respondents are then selected on the basis of referral. Therefore, this procedure is also called referral sampling. Snowball sampling is used in researcher situations where defined target population is rare and unique and compiling a complete list of sampling units is a nearly impossible task.[50] For example, in the case of the earlier discussed example of the manager of brand X of washing machine, if s/he wanted to study the owners of the second hand washing machines it will be very difficult to identify the owners of such washing machines and therefore, snowball sampling may provide a way forward. If traditional probability of nonprobability methods were used for such a study, they will take too much time and incur high costs. The main underlying logic of this method is that rare groups of people tend to form their own unique social circles.[51] While there are several disadvantages in using this procedure as it is a nonprobability technique. However, on the other hand it is a good procedure for identifying and selecting hard-to-reach, unique target populations at a reasonable cost and time.

4.6 Selecting an appropriate sampling technique

As discussed above, both probability and nonprobability sampling techniques have their own advantages and disadvantages. Overall, it depends on various factors to choose the most appropriate sampling technique. A researcher has to consider the research objectives first as to do they call for qualitative or quantitative research. Secondly, available resources should be kept in mind including the time frame available for conducting the researcher and making the findings available. The knowledge regarding the target population as well as the scope or research also is important in selecting the right kind of sampling technique. Researcher should also focus on the need for statistical analysis and degree of accuracy required with regard to the research and the expected outcomes. On the basis of these parameters a researcher can identify an appropriate sampling technique.

4.7 Conclusion

This chapter focused on one of the most important research issue in marketing research, sampling. As detailed in the chapter sampling is quite a common phenomenon in our decision making process. Before delving deeply into the sampling process one must be aware of several basic constructs involved with sampling namely; population, target population, elements, sampling unit and sampling frame. Determining the final sample size for research involves various qualitative and quantitative considerations.

There are two basic techniques of selecting sample; probability sampling techniques and nonprobability sampling techniques. Probability sampling techniques are more robust in comparison to nonprobability sampling. Findings based on nonprobability are hard to generalize to a wider population.

Probability sampling is sub-divided into simple random sampling, systematic sampling, stratified sampling and cluster sampling. While being robust probability sampling techniques are resource intensive in terms of cost and time involved. Nonprobability sampling is sub-divided into convenience sampling, judgement sampling, quota sampling and snowball sampling. Nonprobability sampling techniques are less costly and less time consuming however they have problems relating to selection bias also.

Selecting an appropriate sampling technique depends on various factors such as research objectives, available resources, knowledge of target population and scope of research, degree of accuracy and statistical analysis required for result interpretation.