UNIT 5 SAMPLING

Objectives

After reading this unit, you should be able to:

- describe the importance of sampling in marketing research studies
- define various basic sampling concepts
- explain the process of sampling in marketing studies
- discuss various sampling designs and their applications in marketing research
- estimate the sample size required in a marketing study

Structure

5.1 INTRODUCTION

Sampling plays a vital role in carrying out any marketing research study. There would be hardly any marketing research study which does not involve the use of sampling. It is the backbone of marketing research. A company selling a particular brand of toothpaste and interested in knowing what proportion of households in Delhi use its brand would involve the selection of samples of households. A market potential study aiming at finding the numbers, distribution and socio-economic characteristics of potential customers 'of a product would definetly involve the selection of some kind of sample. A company interested in introducing a new style of packaging for its product would be interested in knowing the reactions of its customers - a study of which would require the selection of some type of sample.

In this unit you will be introduced to various sampling concepts. Census (Complete enumeration), an alternative to sampling would also be discussed. A brief mention of sampling and non-sampling errors will be made. The various probability and non-probability sampling designs as applicable to marketing research will be introduced. An important decision while taking a sample is to know how large a sample should be taken since the choice of sample size involves various elements like time, money accuracy etc. Therefore, the determination of sample size would also be discussed in brief in this unit.

5.2 SAMPLING – ITS IMPORTANCE AND LIMITATIONS

You would hardly come across an individual who is not familiar with sampling. People are used to drawing inferences about a large number of items (called a sample). The following example. The following examples would illustrate as to where the analysis of samples can be used to draw conclusions about the population.

1) Your mother might like you to taste a small quantity of egg pudding to form an opinion about the acceptable quality of the entire dish.
2) You might have noticed at the book-stalls that before buying a book or a magazine, people generally flip through its few pages (a sample) to determine whether it is of interest to them.
3) A manufacturer of electric bulbs interested in estimating the average life span of a bulb would put a few bulbs (a sample) for complete use to achieve the objective.
4) You don't have to drink a whole bottle of squash to say that it tastes good or bad - a small amount should be able to do the job.
5) You go to the market to buy a bag of `wheat flour'. To determine its quality, you take a handful of it and examine its quality to draw conclusions about the quality of the whole bag of wheat flour.

These are just a few examples from everyday life where samples are used to draw inferences about the entire population (universe). An alternative to sampling is complete enumeration (census). Under complete enumeration, we collect data for each and every unit (person, household, factory, shop etc.) belonging to the population, which is the aggregate of all units of a given type under consideration. Sampling has some' major advantages over complete enumeration and therefore it is used very often in marketing research studies. Explained below are some of the advantages of sampling over complete enumeration.

The data collection through sampling definitely involves a lower cost as compared to census. In. a marketing research project aiming to study the buying behaviour of households in Delhi for consumer non-durable items, the cost of interviewing 3000 households would be much less than interviewing all the households in Delhi, Therefore, from a managerial points Of view, a sample's economic efficiency is very attractive.

A sample saves time. Managers have', a time frame in which they are supposed to take decision on the basis of whatever information they can gather during that time period. They cannot wait till the information obtained on the basis of complete enumeration is available. In the above example, if data is to be collected on all the households in Delhi, we would require more time *to* print additional questionnaires, more interviewers will have to be trained and further the data analysis would require more time.

It is not at all essential that the *accuracy* 1of information may be enhanced by taking *a* complete enumeration. *Suppose* we want to estimate the average price of Camps Cola charged by restaurants in Delhi at *a* point of time. As we are aware that prices usually don't vary widely over a cross-section, it would be a wastage of time and money to go for complete enumeration. One could always obtain very reliable estimates by taking a representative; sample.

A sample is better in situations in which measuring of a particular element from a group would destroy the elements or render them useless after examination. Let us consider the example of testing a photographic film. To test the quality of the film, the film has to be exposed which infact destroys it for further use. It is not possible to go for complete enumeration as there would be no film left if all were tested. Therefore, a sample of each batch should be used in a particular production run to judge the quality of the film.

The above discussion should not lead you to conclude that carrying out a census is useless. There are some situations under which a census may be preferable to a sample. A census is desirable when the population is small, the variance in the characteristic being measured is large and the fixed cost of sampling is very high. There are also problems associated with sampling which we would discuss shortly.

Problems Associated with Sampling

Let us consider a hypothetical population of size 10,00,000 households whose monthly food expenditure is of interest to us. If we know the food expenditure of all (10,00,000) households in the universe, we could then compute its arithmetic mean (usually referred to as population mean). Let us assume that the population mean (μ) is Rs.2,500. However, in reality it may not be possible to go for a census for the reasons mentioned earlier. Therefore, one might have to resort to sampling. Suppose a sample of size 1000 is taken, the analysis of which is used to draw inferences about population parameters. Suppose the sample mean (9ˉ: average monthly food expenditure) is Rs. 2,450. We find that though the sample mean is close to the population mean, it is not identical with the true mean (μ) . Suppose we take another sample of size 1000 and compute X. Let the sample mean in this case be Rs.2600. Here also we find a. similar result, that though the sample mean is close to population mean, it is not identical with the true mean: In fact, it is very rare that the sample means would coincide with the true population mean. Why does this happen? There must be certain problems with which you should get acquainted with. We know that the sample data represents only a fraction of population data. Therefore, the problem is related to how well the sample represents the characteristic of the populations of which it is a part. The confidence attached to the sample data is affected by two different types of error called sampling error and non sampling error. If both errors don't exist then in our example, sample mean Y would exactly coincide with population mean for each sample drawn from the population. The two different types of errors viz. Sampling error and non-sampling errors are discussed below:

Sampling Error

The error which arises due to drawing inferences about population parameter on the bais of observations drawn from a sample (a part of the population) is called sampling error. In other words a sampling error is made while selecting a sample which is not representative of the population. It represents the difference between sample value and true value of population parameters. A sampling error is bound to occur while selecting a sample as it is difficult, if not impossible, for a sample (a small part of the population) to be exactly representative of the population. This occurs no matter how careful the researcher is randomly choosing the sample. Sampling error, therefore, is a result of chance. The sampling error usually decreases with increase in sample size and it is non-existent in a complete enumeration survey.

Non-sampling Error

You might expect that a complete enumeration of all the units in the population would result in data free from errors. However, this is not the case. It is difficult to completely avoid errors of observations or ascertainment. While processing the data, one may commit tabulation errors. The main reasons for the occurance of non-sampling errors are the improper specification of the scope of study, inadequate coverage of population and sample, vague definition of variables involved and wrong methods of data collection. A combination of one or more of following factors could always result into non-sampling errors.

1) Inadequate specification of data which may not be consistent with the objectives of survey.
2) Inaccurate methods of interview.
3) Inaccurate reporting by the respondents.
4) Lack of trained and experienced interviewers.
5) Plain lying by the respondents,
6) Inability to locate proper respondents due to improper instructions or wrong address.
7) Inadequate scrutiny of the basic data.
8) Errors in coding, editing and tabulation of data.
9) Errors in presenting the tabulated results, graphs etc

The non-sampling error can occur in the case of complete enumeration as well as in sample survey. However, non-sampling error increases with the increase in sample size. Therefore, the size of non-sampling error is much larger in case of complete enumeration than that of sample surveys. This, is because in a study involving complete enumeration of the units of the population, we need more interviewers, more staff to supervise these interviewers and more manpower to convert raw data to computer input. This would involve more training of the interviewers, which would be both, costly and time consuming. Further, as the size of staff becomes larger their quality falls. It also becomes more difficult to control and supervise their activities. All this would result into more errors And less accurate results thereby contributing to increased non-sampling errors.

Activity 1

List out some marketing research studies carried out in your organisation or the organisation you know of where sample surveys were conducted.

…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
………………………………………………………………………………………..

Activity 2

List out the advantages of sampling over complete enumeration (relate it to the studies in Activity I).

…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
………………………………………………………………………………………..

Activity 3

It is possible to control sampling error? If so, how? -

…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
………………………………………………………………………………………..

Activity 4

Make a list of he sources of non-sampling error. What do you think should be done to control them?

…………………………………………………………………………………………
…………………………………………………………………………………………
…………………………………………………………………………………………
………………………………………………………………………………………..

5.3 SAMPLING: BASIC CONCEPTS- '

Element : An element is the unit about which information is collected. This provides the basis for analysis according to a well defined procedure. It is important that these should be well defined and it is possible to identify them physically. In marketing research studies, the households, family, shop, factory etc. might constitute elements. The type of unit considered in any marketing research study would depend upon the objectives of the study. For instance, a shop may be considered as a unit in a retail stores survey whereas a household may be treated ' as a unit in a family budget enquiry.

Population : Population or universe is the aggregate of all elements possessing certain specified characteristics which the researcher wishes to study and defined prior to the selection of sample. A properly designated population must be defined in terms of elements, sampling units, extent and time. Let us consider the following examples.

1) A survey of consumers might specify the relevant population as:
   a) Element: Males 35-40
   b) Sampling Unit: Males 35-40
   c) Extent: Delhi
   d) Time: March 10-April 30, 1990
2) Another survey of consumers might specify the relevant population as:
   a) Element: Females above 20 years
   b) Sampling Unit: Households, then females over 20 years
   c) Extent: Delhi
   d) Time: April 15-25, 1990
3) For conducting a socio-economic survey in a region, we might specify the relevant population as:
   a) Element : Households
   b) Sampling Unit: Tehsils, Villages; then households
   c) Extent : Amritsar District
   d) Time: 1989

Sampling Units: A sampling unit is that element or elements considered available for selection in some stage of the sampling process.- In single stage sampling, sampling units and the elements are same. For instance, in our first example of specifying the population (see example above), both sampling units and elements were "males 35-40". This is an example of direct single stage sampling process as the sample of males 35-40 is selected directly.

In a more complex situation (see example 2 above), we are selecting females over 20 indirectly through a two stage process. First of all, we select a sample of households. Then, within these selected households, we select a sample of females over 20. It is only at second stage (final stage in this case) where elements and sampling units are identical.

The third example is more complex than even the second one. In this case a sample of households may be selected in three stages. First of all a sample of Tehsils is selected. Then a sample of villages is selected from each selected Tehsils after making a list of all the villages in it. Finally a sample of households is selected from each selected village after listing all the households in it. In this example Tehsils are taken as first stage unit, villages as second stage unit and households as the third or the final stage unit. You may note that it is at the final 'stage that elements and sampling units are identical.

Sampling Frame: It is a list of all sampling units belonging to the population to be studied with their proper identification and available for selection at a stage„of sampling process. In fact, the actual sample is drawn from the sampling frame. Therefore, we should ensure that the sampling frame contains all the sampling units of the population under consideration. It should exclude units of any other population. The sampling frame should be upto date and free from errors of omission and duplication of sampling units. In fact in marketing research studies a lot of time and effort is spent on preparing a suitable sampling frame. The examples of sampling frame are telephone directory, a list of registered voters, naives of all the students registered with a. particular university, a map, an employee list etc.

Usually the availability of sampling frame defines the population. Each stage of sampling process requires- its own sampling frame. For instance, in three stage sampling process (see, example 3 under population), we would need three sampling frames. These are (1) a list of tehsils (2) a list of villages under the selected tehsils (3) a list of households within the selected villages.

Although a lot of attention is paid to preparing a suitable sampling frame, most of them are far from being perfect.` For example, if we are conducting a study which involves surveying the households of Delhi by telephone, the most obvious sample frame which would come to our mind is the use of telephone directly. However, this may not be an appropriate frame as all the households in Delhi may not have telephone connection for one reason or the other. Further; some numbers may ?hot be listed in the telephone directory which others are listed may not longer be residents of Delhi. Therefore, the survey would not be free from the inaccuracies of population's frame leading to frame's error. In reality, the sampling frame of most of marketing research studies have one or the other shortcomings.

Study Population : A study population is the aggregation of elements from which the sample is actually drawn. You might recall that previously population was defined as the aggregate of elements (possessing certain characteristics) prior to the selection of sample". Because of certain unavoidable problems, the actual sample is selected from somewhat different population from the one defined prior to the selection of sample. This is because it is very seldom that every element which satisfies our definition of a population actually has a chance of being selected. Our list may be incomplete as some elements are likely to be omitted from a list of population because of certain reasons such as some people may have unlisted phone numbers, a. map may not include a new street, a list of registered voters may be incomplete.

Therefore, the study population is the aggregation of elements from which the sample is actually drawn and it is with reference to this (study) population that the inferences are drawn.

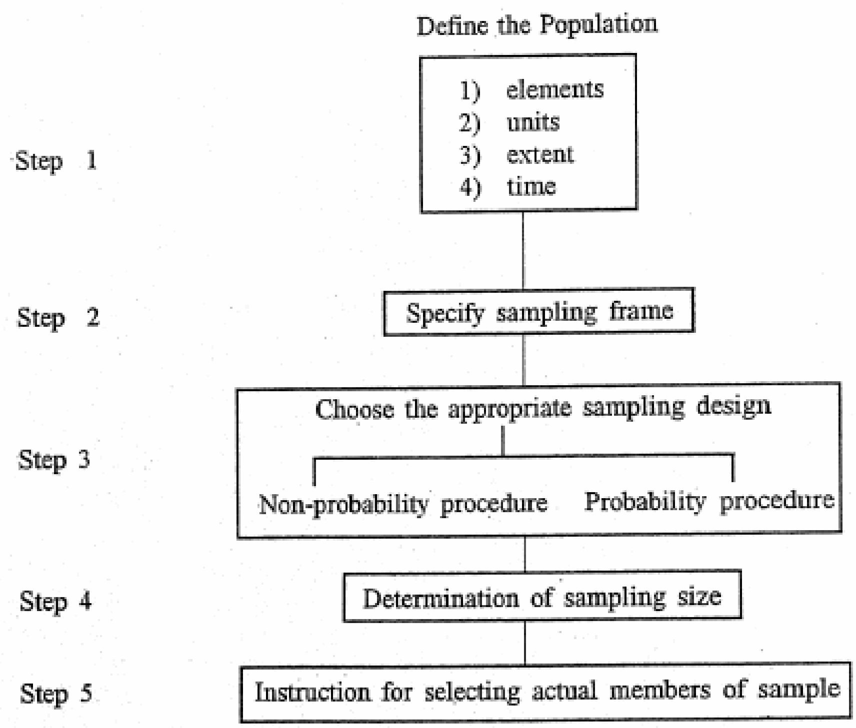
Activity 5

Make a list of few marketing research studies undertaken in your oganisation or an organisation you know of where the elements of population were selected directly (single stage process) and indirectly (two stage or three stage or multistage process).

....
....
....
....

5.4 STEPS IN THE SAMPLING PROCESS

The sampling process consists of five sequential steps. These steps are listed and briefly explained below. An overview of the various phases of sampling is shown in the Figure below.



Step 1: The primary step is to define the population. It should be defined in terms of sampling (i) the elements, (ii) the sampling units, (iii) the extent, and (iv) the time.

Step 2 : The next step is to specify the sampling frame. It has already been explained - in detail in the previous section..

Step 3: The third step is to choose an appropriate sampling design for selecting the same .The sampling design describes the procedure by which a sample is selected. There are two types of procedures namely non-probability sampling procedures and probability sampling procedures. Under probability sampling procedure, each element has a known chance of being selected in the sample. In the non-probability sampling procedure, there is no known chance of an element of the population being selected in the sample. The selection of an element of population in the sample depends upon judgement of the researcher or of the field interviewer. A detailed discussion on this will be taken nip in the next section.

Step 4 : Once we have decided the procedure by which sample would be selected, the next step is to determine how large sample should be taken. A detailed discussion on the determination of size of the sample will find a place in section 5.6.

Step 5: The interviewers who are to go to field to collect actual data need very clear and accurate instructions as to how to do this job. The sampling planning is therefore not complete until these instructions are prepared and handed over to the investigators.

5.5 SAMPLING DESIGNS

The sampling designs describe the procedure by which sample is selected. There exist two classes of methods by which samples can be selected.They are probability sampling methods and non-probability sampling methods.

Probability Sampling Methods : In probability sampling methods every element of the population has a known chance of being selected. Please note that the term known chance does not mean equal chance. Equal chance probability sampling is a special case of probability sampling, called simple random sampling. In probability sampling methods, there is no chance of arbitrary or biased selection and therefore the laws of probability apply. Therefore, it permits us to measure the sampling error which is the difference between the population value and the sample, value.

There are a number of different sampling procedures that fall under probability sampling methods. Some of these methods are listed as under:

1) Simple Random Sampling
2) Systematic Sampling
3) Stratified Sampling
4) Cluster Sampling

We will soon take up the discussion on the above methods.

Non-Probability Sampling Methods: These methods do not provide every item of population any known chance of being selected in, the sample. Here there is no attempt to select a representative sample. The elements of samples are selected on the convenience and/or judgement of the researcher or field interviewer. The selection process is subjective. Since the °sample is not representative of population, it is not possible to make an estimate of sampling error. Also,,we cannot say whether our sample estimates are correct or not.

There are many different sampling methods which fall under the umbrella of non-probability sampling methods. These are listed below:

1) Convenience sampling
2) Judgement sampling
3) Quota sampling

A discussion on these sampling methods will be taken up soon.

It may be worth mentioning [hat most of the marketing research studies make use of non-probability methods of sampling. Since ender these methods, the selection process of these samples are subjective, therefore one should not conclude that the results obtained from it are inferior to what one would obtain by using probability sampling methods. Also, the sample obtained through such methods need not be less representative of the population. For example, a. marketing researcher is asked to develop an index of performance of sales force by measuring items such as sales per salesman; number of calls per day, order call ratio, number of customer complaints and so on. Any particular item is included in the list because the market researcher feels that is represents 'performance'. His non-probability method of selection is a better way to achieve a representation of the population than by a random selection of items representing various characteristics of performances.

We would now discuss some of the non-probability sample selection methods that are used in marketing research studies.

Non-Probability Sampling Methods

1) **Convenience Sampling:** Under convenience sampling, as the name implies, the samples are selected at the convenience of the researcher or investigator. Here, we have no way of determining the representativeness of the sample. This results into biased estimates. Therefore, it is not possible to make an estimate of sampling error as the difference between sample estimate and population parameter is unknown, both in terms of magnitude and direction. It is therefore suggested that convenience sampling should not be used in both, descriptive and causal studies as it is not possible to make any definitive statements about the results from such a sample.

This method may be quite useful in exploratory designs as a basis for generating hypotheses. The method is also useful in testing of questionnaire etc. at the pretest phase of the suudy. Convenience sampling is extensively used in marketing studies and otherwise. This would be clear from the following examples.

i) Suppose a marketing research study aims at estimating the proportion ` of Pan (beetle leave) shops in Delhi which store a particular drinksay Maaza. It is decided to take a sample size of 100. What investigator does is to visit 100 pan shops near his place of residence as it is very convenient to him and observe whether a Pan shop stores maaza or not. This definitely is not a representative sample as most pan shops in Delhi had no chance of being selected. It is only those pan shops which were near the residence of the investigator that had a chance of being selected.

ii) The other example where convenience sampling is often used is in test marketing. There might be some cities whose demographic make tips are approximately the same as national average. While conducting marketing tests for new products; the researcher may take samples of consumers from such cities and obtain `consumer evaluations' about these products as these are supposed to represent "national" tastes.

iii) A ball pen manufacturing company is interested in knowing the opinions about the ball pen (like smooth flow of ink, resistance to breakage of the cover etc.) it is presently manufacturing with a view to modify it to suit customers needs. The job is given to a marketing researcher who visits a college near his place of residence and asks a few students (a convenient sample) their opinion about the ball pen in question.

iv) As another example, a researcher might visit a few shops to observe what brand of vegetable oil people are buying so as to make inference about the share of a particular brand he is interested in.

2) Judgement Sampling: Judgement sampling is also called purposive sampling. Under this sampling procedure, a researcher deliberately or purposively draws a sample from the population which he thinks is a representative of the population. Needless to mention, all members of the population are not given chance to be selected in the sample. The personal bias of the investigator has a great chance of entering the sample and if the investigator chooses a sample to give results which favours his view point, the entire sutdy may be vitiated.

However, if personal biases are avoided, then the relevant experience and the acquaintance of the investigator with the population may help to choose a relatively representative sample from the population. It is not possible to make an estimate of sampling error as we cannot determine how, precise our sample estimates are.

Judgement sampling is used in a number of cases, some of which are mentioned below:

i) Suppose we have a panel of experts to decide about the launching of a. new product in the next year. If for some reason or the other, a member drops out from the panel, the chairman of the panel may suggest the name of another person whom he thinks has the same expertise and experience to be a member of the said panel. This new member was chosen delibertely - a case of Judgment sampling.

ii) The method could be used in a study involving the performance of salesmen. The salesmen could be grouped into top-grade and low-grade performer according to certain specified qualities. Having done so, the salesmanager may indicate who in his opinion would fall into which category. Needless to mention, this is a biased method. However, in the absence of any objective data, one might have to resort to this type of sampling.

3) Quota Sampling: This is one of the most commonly used sampling method in marketing research studies. Here the sample is selected of the basis of certain basic parameters such as age, sex, income and occupation that describe the nature of a population so as to make it representative of the population. The investigators or field workers are instructed to choose a sample that conforms to these parameters. The field workers are assigned quotas of the numbers of units satisfying the required characteristics on which data should be collected. However, before collecting data on these units the investigators are supposed to verify that the units qualify these characteristics.

Suppose we are conducting a survey to study the buying behaviour of a product and it is believed that the buying behaviour is greatly influenced by the income level of the consumers. We assume that it is Possible to divide our population into three income strata such as high income group, middle income group and low income group. Further, it is known that 20%0 of the population is in. high income group, 35% in the middle income group and 45% in the love income group. Suppose it is decided to select a sample of size 200 from the population. Therefore, samples of size 40, 70 and 90 should come from high income, middle income and low income groups respectively. Now the various field workers are assigned quotas to select the sample from each group in such a way that a total sample of 200 is selected in the same proportion as mentioned above. For example, the first field worker may be assigned a quota of 10 consumers from the high income group, 25 from the middle income group and 40 from the low income group. Similarly the 2nd field worker may be given a different quota and so on such that a total sample of 200 is obtained in the same proportion as discussed earlier.

The above example was very simple one, however, suppose we are told further that the buying behaviour is not only influenced by his income but also by his age (categorised as 45 and above or below 45). With this additional character, suppose the distribution of population (universe) is as follows:

Table 1 . Distribution of Population (%)

Income / Age	High Income	Middle Income	Low Income	Total
45 and above	12	10	30	52
below 45	8	25	15	48
Total	20	35	45	100

The above table indicates that in this universe there are 12% of people in the high income group and fall in the age group of 45 and above, there are 25% people in middle income group and below the age of 45 and so on. Suppose, it is decided to take a sample of size 200. Therefore, the distribution of the sample conforming to these two parameters (in the same proportion as population) would be as follows:

Table 2: Distribution of Sample (Numbers)

Income / Age	High Income	Middle Iucorne	Low Income	Total
45 and above	24	20	*60*	104
below 45	*16*	50	30	96
Total	4U	70	90	200

The above table shows that a sample of 30 should be taken from the population with low income and below the age of 45. Similarly a sample of 20 should be taken from the population with middle income and having age of 45 and above, and so on.

Now, having decided the size of sample falling under each of the six cells ("high income and below 45", "middle income with 45 and above" and so on), we fix the quotas for each of the field worker to collect data conforming to the above norms so as to obtain a total sample of size 200.

At the outset, the Quota sampling procedure might look similar to stratified sampling (to be discussed under probability sampling designs). However, there is a difference between the two. Under stratified sampling, the, field-Worker selects a random sample from each cell of the population, whereas under Quota sampling the selection of sample is not random. It is left to the judgement of ti>e field worker.

The Quota sampling method has some weaknesses. These are listed below:

i) It is usually difficult to obtain an accurate and up to date proportion of respondents assigned to each cell.

ii) As the number of parameters (control characteristics) associated with the objectives of the study become large, the total number of cells increase. This makes the task of field staff difficult as it may not be easy to get a. desired respondent.

iii) It is very, important that all of the proper parameters (control characteristics) related to the study in question must be incorporated while taking sample. The results of the study could be misleading if any relevant parameter is omitted for one reason or the other.

iv) The field workers might like to visit those areas where the chances of the availability of a respondent satisfying certain desired parameters is very high. Further, the field workers might avoid certain respondents who look unfriendly and live in houses which may not be of good appearance. These factors are likely to make the findings of the study less reliable.

The Quota sampling method has some advantages too. The method has a lower cost " and field workers have a free hand to select respondents for each cell to fill their quota. The samples, if selected with care would result into more definitive findings.

Probability Sampling Methods

1) Simple Random Sampling: Under this sampling design, each member of the population has known and equal probability of being included in the sample. For details on now to draw samples using this sampling procedure, please review unit 13 of MS-8 course. Simple random sampling is not widely used in marketing research because of the following reasons.

i) In consumer research studies, we usually select individuals, households, shops or areas as the sampling units. It may not be easy to prepare a sampling frame as it is very difficult to get lists of households, individuals and shops, although areas may be completely represented through maps.

ii) We know that an industry comprises of various firms of different sizes. If one wants to study some aspects of an industry, one might like to choose a sampling design where there is a higher probability of a larger firm being selected. If that is the case, the very concept of simple random sampling becomes inapplicable in such situations. The simple random sampling has some applications in Industrial Marketing where generally purchasing agents or companies or areas are the sampling units which are usually not very big in number. Therefore, it becomes easy to prepare a sampling frame thus facilitating the use of simple random sampling.

2) Systematic Sampling: The mechanics of taking a systematic sample are very simple. Suppose the population consists of ordered N units (numbered from 1 to N) and a sample of size n is selected from the population in such a way that $\frac{N}{n} = K$ (rounded to the. nearest integer). Here K is called a sample interval. Systematic sampling then consists in selecting a number at random between 1 and K (both inclusive) and then selecting every subsequent Kth unit till a sample of size n is obtained.

To make the above more clear, let us assume that we have an ordered population of size N=500. Suppose it is decided to take a sample a size n=50. Therefore, our sampling interval would be $\frac{N}{n} = \frac{500}{50} = 10$ We then select a number at random between 1 to 10 (both inclusive) Suppose it turns out to be 6. Then our sample units would be 6,16, 26, 36 and so on.

Systematic sampling is a case of mixed sampling where both probabilistic and non-probabilistic methods of choosing a sample are used. This is because the first unit of the sample is selected at random between numbers 1 and K (probabilistic method) and then the rests of the units of the sample are fixed by the choice of the first member (non-probabilistic method).

It is very likely that systematic sampling would result into more representative sample than simple random sampling. In systematic sampling the elements of the population are ordered in a particular fashion. Suppose we want to estimate the sales of all the retail stores in Delhi. Under a simple random sampling, if we draw a random sample of size n, it is very likely that Most of the sampled stores might turn out to be low sales volume store. However, in systematic sampling we order these retail stores according to ascending or descending order of sales, therefore, a systematic sample would definitely contain some low volume and high volume retail stores. Thus, a systematic sample is likely to be more representative than a sample random sample.

A systematic sample might also reduce the representativeness of the sample. This could happen if items are ordered in such a way to produce a cyclical pattern. Suppose a supermarket is interested in estimating average daily Sales by using a sample of certain daily sales. Suppose a sampling interval of seven days is chosen.. The sample, would therefore, result in recording of the sales on the same day of the week. This would not reflect day-of-the week variations in sales. For example, a sample of Sundays would generally overstate sales and a sample of Wednesdays and understate sales.

3) Stratified Sampling: In stratified sampling, the entire population is divided into various mutually exclusive and collectively exhaustive strata (groups). By mutually, exclusive it is meant that if an element of a group belongs to one strata, then it doesn't belong to any other strata. By collectively exhaustive we mean that all the elements of various strata put together completely cover all the elements of the population. The groups (strata) are created on the basis of a variable (criteria) known to be correlated with the variable under study. The possible criteria for stratification of a population could be income of the individuals, age, sex, frequency of a purchase of a product, size of the household, size of the retail store, region of the country mid so on. The stratification is also possible on the basis of more than one variable. This, of course, increases the number of stratum. :he cost of stratification may come as a constraint in increasing the number of grout_ (stratum). A variable which is considered to be good in stratification of one population may not be so in the case of other. However, one thing should be kept in mind that stratification should be done in such a way so as to minimise the variability among sampling unit within strata (more homogeneous) and maximise the variability among strata (more hetrogeneous).

Once the population has been divided into various strata, separate simple random sample of various sizes are selected from each stratum. There might be cases, where systematic or any other type of simple random samples may be selected from each stratum and the resulting design may still be called stratified sampling. The selection of the size of sample from each stratum can be done on either proportionate or disproportionate basis. In proportionate stratified sampling, the number of members selected from each stratum is proportional to its share of the total population. However, in case of disproportionate stratified sampling, the number of members selected from each striatum is not proportional to its share of total population, The choice of proportionate or disproportionate sampling method among strata depends upon whether the variances in each group (stratum) are equal or not. If the variance of each stratum is almost equal, one should go for proportionate stratified sampling. If variances are not equal, a large sample should be taken from the stratum with large variance.

Stratified sampling is moderately used in marketing research. Suppose we are interested in estimating the retail sales of Kwality, (a brand name) .tea in Delhi. Before using stratified sampling method in its estimation, we may ask the question "What factors account for the variations in the retail sales of Tea?" We may possibly get the following two answers.

i) Size of the store-whether it is large, medium or small, Large stores are supposed to sell more tea than small stores,

ii) Day of the week, the sales is supposed to be more during week ends than during the week' days.

We have now three types of stores and two types of days of the week. So the total number of stratum (Groups) on which the total population can be divided is si^{p}:. They are "large store and week days, *medium* store and weekend' arid sc~ on". Once we are able to make six stratum, the sample front each stratum can be selected either according to proportionate stratified sampling scheme or- disproportionate stratified sampling scheme depending upon the variability in each stratum, One can similarly think of many other marketing research studies where stratified sampling method could be used.

4) Cluster Sampling: If we divide all the elements of the population into suitable, clusters; and select few clusters randomly and all the elements of the selected clusters are used, then this method of sampling is called cluster sampling'. This method of collecting data is cheaper since collection of data fear nearby units is easier, faster and more convenient than collecting data over units scattered over a region. For instance, it would not only be cheaper but also convenient to collect data on all households in a sample of few villages.(clusters) than to surrey a sample of the same number of households selected randomly from a list of all households.

The criteria for dividing the population into mutually exclusive and collectively exhaustive clusters is, that the elements in the clusters should be as heterogeneous as possible and elements between cluster should be as homogeneous as possible.

To make the concept of cluster sampling very clear let us consider a hypothetical population consisting of 24 elements divided into four equal-sized clusters as shown below:

Cluster	Population element number
Cluster 1	1, 2, 3, 4, 5, 6
Cluster 2	7,8 , 9, 10, 11,12
Cluster 3	13, 14, 15, 16, 17, 18
Cluster 4	19, 20, 21, 22, 23, 24.

Suppose we wanted to select a probability sample of size '12. We could do it in two ways. We could either take a simple random sample of size 12 from the population or alternatively we could select at random two clusters out of four clusters and use all the elements of these selected clusters. In both simple random sampling and cluster sampling, the sample size as a fraction of population size is same. However, all possible combination of elements are not equally likely in cluster sampling. Many of the combinations would be impossible.

5) Area Sampling

In a marketing research study involving sampling of population which may be grouped according to geographical areas (blocks), Census tracts, Communities, constituencies etc., another version of cluster sampling namely Area Sampling is used.

The entire area is divided into various clusters. The cluster may or may not be of equal size. Below we will discuss a sampling scheme where sampling is done by taking into account the size of the cluster. This type of design is called probability proportional to size sampling.

Probability Proportional to Size: This sampling design is used when we have to sample cluster of varying sizes. Suppose there is a tome divided into seven blocks (clusters) of varying sizes. Assume that in one of the blocks the size of the household is 120 whereas in the other block the size of the household is 80. The block having 120 size household should have greater probability of being selected than the one with a size of 80. In fact the probability in the first case should be 1.5 times the probability in the second case. This concept is used in this sampling design so that a greater weightage is given to the cluster with higher size.

Let us consider the data given in the following table.

Block No.	Number of household	Cumulative No. of households	Associated random numbers
1.	120	120	001-120
2.	80	200	121-200
3.	210	410	201410
4.	190	600	411-600
5•	350	950	601-950
6.	150	1100	951-1100
7.	400	1500	1101-1500
	1500		

In the above table a small area has been divided into seven blocks. It is evident from the above table that the number of households in each block are varying. If a sample of size 30 is to be selected from a population of 1500, each household should have a probability of 30 - 1500 = .02 being selected in the sample. We have identified a total of 1500 household in seven blocks (clusters). We will assign a number from l to 1500 for each household. This is shown in the last coulmn of the above table. Suppose we have to select at random three clusters. We will look at the four digited random number from the random number tables with numbers 0001 to 1500. We select three numbers. Suppose the numbers are 476; 1253 & 129. This means cluster numbering 4, 7 and 2 are selected. Now we would select a sample of size 10 from each of the clusters. The probability of selecting the required household are as follows:

Probability of household in block A = Number of blocks to be chosen X block probability X within block household probability

For household in block 2 = $3 \times \frac{80}{1500} \times \frac{10}{80}$ = 02

For household in block 4 = $3 \times \frac{210}{1500} \times \frac{10}{210}$ = 02

For household in block 7 = $3 \times \frac{400}{1500} \times \frac{10}{400}$ = 02

We find that irrespective of the block size, the probability of selecting a household equals .02. This is the same as required by the over all sampling design.

We note that in the probability proportional to size (PPS) design the blocks with larger size are given more weightage selected in the sample. It can be shown that the efficiency of the estimate increases by this procedure as compared to when all the blocks (cluster) have equal probability of being selected.

Activity 6

Make a list of some of the marketing research studies where some of the non-probability sampling designs and probability sampling designs could be used. Also justify the choice of a particular sampling design you have selected for a. study.

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Activity 7

Why higher weightage is given to cluster with higher size?

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

5.6 ESTIMATION OF SAMPLE SIZE

We will discuss below two procedures for estimating sample sizes. The first one is called confidence interval approach which is based on the idea of constructing confidence intervals around sample means or proportions. The second approach is called the hypothesis-testing approach and makes use of both type I and type II errors. The students are advised to revise units 14 and 15 of MS-8 course before reading this section. Let us discuss below the two approaches.

1) The Confidence-Interval Approach

In unit 14 of MS-8, we have discussed the procedure for constructing confidence intervals for population mean and population proportion. While doing so, we used the following notations:

μ = population mean

$\bar{x}$ = sample mean

σ = standard deviation of the population

s = standard deviation of the sample

$\bar{p}$ = sample proportion

p = population proportion

n = number of items in the sample (sample size)

Suppose we want to use the mean of a random sample to estimate the mean of a population and we want to be able to assert with a (1-a) % confidence that the allowable error of this estimate is e. The question is how large a sample need to be taken to achieve this?

We know that $Z = \dfrac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$

Where Z is a standard normal variate and n is the required sample size. The quantity $\bar{x}$ - μ represents allowable error e.

There fore we have

$$z = \frac{e}{\frac{\sigma}{\sqrt{n}}}$$

$$z = \frac{\sqrt{n}e}{\sigma}$$

$$\therefore n = \frac{(z\sigma)^2}{e^2} \qquad ......(1)$$

The value of Z, for a given confidence level can be read from standard normal table. For a 95%o confidence level, the standard normal table indicates that the Z value that allows a 0.025 probability that the population mean would fall outside one end of the interval is Z⁼1.96. Since we are allowing a total probability of 0.05 that the population mean will lie outside either end of the intervals, Z⁼1.96 is the correct value for a 95% confidence level. Similarly Z=2.58 is the correct value for a 99% confidence level.

Let us see how can we use formula (1) in estimating sample size for the following problem.

In a study of television viewing habits, it is desired to estimate the average number of hours that teenagers spend watching per week. Suppose the population standard deviation is given to be 192 minutes. How large a sample would be required if one wants to be able to assert with 95% confidence that the sample mean would *differ* from the population mean by at most 24 minutes?

Here e = 24

σ = 192

Z = 1.96 (for a 95% confidence level)

$$n = \frac{(z\sigma)^2}{e^2}$$

$$= \frac{(192 \text{ x } 1.96)^2}{24^2}$$

= 245.86

= 246 (rounded to the next integer)

Data Collection Similarly if we want to use sample proportion (p) to estimate the population proportion (p) and we want to be able to assert with a (1-(x)% confidence that the allowable error of this estimate is e. Again we ask the question: How large a sample is required to achieve this?

we know that

$$Z = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Where Z is a standa.-4 normal variate and n is the required sample size. The quantity p - p represents allowable error e.

Therefore, we have

$$Z = \frac{e}{\sqrt{\frac{p(1-p)}{n}}}$$

$$Z = \frac{\sqrt{n}e}{\sqrt{p(1-p)}}$$

$$\therefore n = \frac{p(1-p)Z^2}{e^2} \quad \text{...(2)}$$

The population p may be unknown. Therefore, in the above formula (2), we substitute the maximum possible value of p (1-p) which can be shown to be equal to $\frac{1}{4}$ and occurs when p = $\frac{1}{2}$. Therefore, in formula (2), we substitute $p(1-p) = \frac{1}{4}$ and this would increase our sample size and would ensure that the allowable error lies within the prescribed limit and at given confidence level. Therefore, the formula (2) can be rewritten as:

$$n = \frac{1}{4}\frac{Z^2}{e^2} \quad \text{....(3)}$$

Let us explain its use with the help of following example:

An automobile insurance company wants to estimate from a sample what proportion of its thousands of policy holders intend to buy a new car within the next one year. How large a sample is required to be able to assert with 95% confidence that the sample proportion and true proportion will differ by less than 0.03?

Here e = 0.03

Z = 1.96 (fora 95% confidence level)

(The Sample Size) n $= \frac{1}{4}\ \frac{Z^2}{e^2}$

$= \frac{1}{4}\ \frac{(1.96)^2}{(0.03)^2}$

= 1067.11

=1068 (rounded to next integer)

2) The Hypothesis-Testing Approach

The estimation of sample size based on the hypothesis-testing approach makes use of Type I error and Type I1 error. Type I error occurs when we reject H_o: the null hypothesis, when it is true. Type II error occurs when we accept $H_.$ when it is false. We will use a to denote the probability of committing a Type I error whereas p is used to denote the probability of making a Type II error. [For details please revise unit 15 of MS-8 course]

We will discuss the estimation of simple size with the help of a numerical example. Let us first of all consider the case of means of means then that of proportions.

The Case of Means

A large retailer wants to determine whether the ^{p}aean income of families living within two miles of a proposed building site exceeds Rs. 6500. It is given that the population standard deviation is *Rs.* 500. How large a sample will be required if the probability of a Type I error is to be 0.05 and the probability of a Type II error is to be 0.01 when mean income of population is Rs. 6900? To work out this problem, we may note the following:

i) Here, our null (H_0) and alternative (H_1) hypotheses to be tested in terms of population means, μ_0 and μ_1 respectively are as follows:

$H_0 : \mu_0 = 6500$
$H_1 : \mu_1 = 6900$

ii) We are given

α (probability of Type I error) = 0.05
β (probability of Type II error) = 0.01

iii) For a one tailed test the Z values for the 0.05 and 0.01 risks are found from standard normal tables to be $Z_\alpha = 1.64$ and $Z_\beta = 2.33$

iv) The population standard deviation (σ) = 500. Now we have to calculate the sample size that will meet the α and β requirements. For that; we need to solve the following two simultaneous equations :

$$\text{Critical Value} = \mu_0 + Z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$\text{Critical Value} = \mu_1 - Z_\beta \frac{\sigma}{\sqrt{n}}$$

By equating these two equations and solving n, we get

$$n = \frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{(\mu_1 - \mu_0)^2}$$

For our example $Z_\alpha = 1.64$, $Z_\beta = 2.33$, $\sigma = 500$, $\mu_0 = 6500$ and $\mu_1 = 7000$

Therefore, substituting these values in the above formula we get :

$$n = \frac{(1.64 + 2.33)^2 \text{x} (500)^2}{(6900 - 6500)^2}$$

$$= \frac{(15.7609)\,(250000)}{160000}$$

$$= 24.626$$

$$\cong 25$$

The Case of proportion: To discuss the case of estimation of sample size, let us consider the following example:

It has been claimed that 40 percent of all shoppers can identify a highly advertised trade mark . How large a sample will be required if the probability of a type I error is to be trade 0.05 and the probability of a type II error is to be 0.01 when the population proportion is 41%?

we set up the hypotheses

$H_0 : P_0 = 0.40$

$H_1 : P_1 = 0.41$

the value

$Z_\alpha = Z_{0.05} = 1.64$

$Z_\beta = Z_{0.01} = 2.33$

In ordere to estimate the sample size,we use the following formula

$$n = \left[\frac{Z_\alpha \sqrt{P_0(1-p_0)} + Z_\beta \sqrt{p_1(1-p_1)}}{P_1 - p_0} \right]^2$$

$$= \left[\frac{1.64\sqrt{0.4(0.6)} + 2.33\sqrt{0.41(0.59}}{0.01} \right]^2$$

$$= \left[\frac{0.803 + 0.564}{0.01} \right]^2$$

$$= \left[\frac{1.367}{0.01} \right]^2$$

$$= 18686.89$$

$$= 18687$$

After having decided upon the size of the sample, the field workers should be given clear and accurate instruction for collecting data from the field

Activity 8

Suppose we want to estimate what proportion of gift items purchased at department store are returned for a refund ,and we want to assert it with 95% confidence that the error of our estimate will be less than 0.05. How large a sample should be taken?

....
………………………………………………………………………………………
………………………………………………………………………………………
………………………………………………………………………………………
………………………………………………………………………………………

Activity 9

The personnel director of a manufacturing company wants to estimate the average mechanical aptitude (as measured by a certain test) of a large group of employees, and she wants this estimate to be in error by at most 2.0 with 99% confidence level. If it is presumed from experience that population standard deviation n= 15.0 for this test, what should the sample size be?

………………………………………………………………………………………
………………………………………………………………………………………
………………………………………………………………………………………
………………………………………………………………………………………
………………………………………………………………………………………

Activity 10

Suppose we want to test the null hypothesis μ =Rs.400 against the alternative μ≠ Rs 400 for a population whose standard deviation is rs. 12. If this hypothesis is true, we want to be 95% confident of accepting it, and if the true mean is Rs. 405, we want to be 90% sure of rejecting the hypothesis. What is the required sample size?

……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………
……………………………………………………………………………………………

Activity 11

If one sets up the hypotheses

$H_{o:}\ p_1 = 0.4$

$H_l : p_1 = 0.5$

With a risk of 0.1 and a beta risk of 0.05 if $p_1 = 0.7$, what is the appropriate sample size under the hypothesis testing approach?

5.7 SUMMARY

In this unit, we have highlighted the importance of sampling in marketing research studies. The advantages of sampling over the complete enumeration are explained. The unit also points out the problems associated with sampling and discusses why sampling estimate may deviate from the population estimate (true value). In this connection, a brief discussion on both sampling and non-sampling error finds a place in this unit.

The various steps in the sampling process are also briefly explained. The two classes of sampling methods namely probability sampling methods and non-probability sampling methods are brought out in this unit. Under non-probability methods - the sampling designs like convenience sampling, judgement sampling and quota sampling have been included. The sampling designs like simple reandom sampling, systematic sampling, stratified sampling and cluster sampling find a place under probability sampling methods. The determination of sample size by (i) the confidence - interval approach, and (ii) the hypothesis testing approach forms the concluding section of the unit.

5.8 SELF-ASSESSMENT QUESTIONS

1) Compare and contrast quota sampling with stratified sampling.

2) Point out the basic difference between stratified sampling and cluster sampling.

3) List out the possible sources of errors in the following procedures:

   i) A basket of apples is sampled by taking some apples from the top.

   ii) Investigators collecting data for a study on household expenditure in a town conduct house-to-house enquiry of the households selected at random, during the working hours of the day ignoring those houses from which there is no reply.

   iii) For a study aimed at estimating the daily average sale of a supermarket, an investigator collects sales figures on every Tuesday of the week for a period of six months.

   iv) To get the feedback on the new T.V. serial presently being telecast on Doordarshan, the producer of the serial collects the opinion of the people living in his neighbourhood.

4) Explain in detail the various sampling designs under non-probability sampling method. Also bring out their relevance in marketing research studies. .

5) Data Collection 5) Define stratified random sampling. What stratification criteria would you employ in sampling public opinion on TV serial `Mahabharat'?

6) It is desired to estimate the mean lifetime of a certain kind of vacuum cleaner. Given that population standard deviation 6 = 320 days, how large a sample is needed to be able to assert with a confidence level of 99% that the mean of the sample will differ from the mean of the population by less than 45 days.

7) Suppose we want to estimate what percentage of the drivers are actually exceeding the 50 kmph speed limit on a stretch of road between Moti Nagar and Vikas Puri in New Delhi. How large a sample would be required to be able to assert with a 95% confidence level that the sample proportion and the true proportion will differ by less than 4 per cent'?

5.9 FURTHER READINGS

Freund, John E. and Frank J. Williams, "Elementary Business Statistics - The Modern Approach", Prentice Hall International Editions.

Kinnear, Thomas C. and James R. Taylor, "Marketing Research - An Applied Approach" McGraw-Hill International Editions.

Luck, David J. and Ronald S. Rubin, "Marketing Research" Prentice Hall of India M. Ltd..

Green, Paul E. and Donald S. Tull "Research for Marketing Decisions" Prentice Hall of India Pvt. Ltd.

Westfall, Boyd and Stasch, "Marketing Research - Text and Cases" Richard D. Irwin, Inc.