UNIT 7: SAMPLING

UNIT STRUCTURE

7.1 Learning Objectives

7.2 Introduction

7.3 Sampling Concepts

7.4 Sampling Error and Non Sampling Error

7.5 Sampling Design Process

7.6 Non- Probabilistic and Probabilistic Sampling Technique

7.6.1 Non- Probabilistic Sampling Technique

7.6.2 Probabilistic Sampling Technique

7.7 Determination of Sample Size

7.8 Let Us Sum Up

7.9 Further Reading

7.10 Answers to Check your progress

7.11 Model Questions

7.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- differentiate a sample from a census and identify the conditions that favour the use of a sample versus a census.
- discuss the sampling design process
- classify sampling techniques as non probability and probability sampling techniques.
- describe the non probability sampling techniques of convenience, judgmental, quota, and snowball sampling.
- describe the probability sampling techniques of simple random, systematic, stratified, and cluster sampling.

- learn the determination of the Sample Size in a Research Project.
- discuss the Means and Proportions method for determination of Sample Size.

7.2 INTRODUCTION

This chapter introduces the fundamental concepts of sampling and the qualitative considerations necessary to answer these questions. We address the question of whether or not to sample and describe the steps involved in sampling. Next, we present non probability and probability sampling techniques. We discuss the method of how to determine the sample size in the research process.

7.3 SAMPLING CONCEPTS

Sampling is one of the components of a research design. The formulation of the research design is the third step of the marketing research process. At this stage, the information needed to address the marketing research problem has been identified and the nature of the research design (exploratory, descriptive, or causal) has been determined. Furthermore, the scaling and measurement procedures have been specified, and the questionnaire has been designed. The next step is to design suitable sampling procedures. Sampling design involves several basic questions:

(1) Should a sample be taken? (2) If so, what process should be followed? (3) What kind of sample should be taken? And (4) How large should it be?

7.4 SAMPLING ERROR AND NON-SAMPLING ERROR

A **population** is the total of all the elements that share some common set of characteristics. Each marketing research project has a uniquely defined population. The objective of most marketing research projects is to obtain information about the characteristics of a population, for example, the proportion of consumers loyal to a particular brand of toothpaste. The researcher can obtain information about population characteristics by taking either a census or a sample. A **census** involves a complete count of all

elements in a population. A **sample** is a subgroup of the population selected for the marketing research project.

Budget and time limits are obvious constraints favouring the use of a sample. A census is both costly and time-consuming to conduct. A census is unrealistic if the population is large, as it is for most consumer products. In the case of many industrial products, however, the population is small, thus making a census feasible as well as desirable. For example, in investigating the use of certain machine tools by automobile manufacturers, a census would be preferred to a sample. Another reason for preferring a census in this case is that variance, or variability, in the characteristic of interest is large. For example, machine tool usage at Ford will vary greatly from the usage at Honda. Small population sizes as well as high variance in the characteristic to be measured favour a census.

If the cost of sampling errors is high (e.g., if the sample omitted a major manufacturer like Ford, the results could be misleading), a census, which eliminates such errors, is desirable. **Sampling errors** are errors resulting from the particular sample selected being an imperfect representation of the population of interest. **Non Sampling errors,** on the other hand, can be attributed to sources other than sampling. They result from a variety of causes, including errors in problem definition, approach, scaling, questionnaire design, survey methods, interviewing techniques, and data preparation and analysis. The high cost of non sampling errors (e.g., interviewing errors because of lack of supervision), on the other hand, would favour sampling. A census can greatly increase non sampling errors to the point that these errors exceed the sampling errors of a sample. Non sampling errors are found to be the major contributor to total error, whereas random sampling errors have been relatively small in magnitude. Hence, in most cases involving large populations, accuracy considerations would favour a sample over a census. This is one of the reasons that the Bureau of the Census checks the accuracy of various censuses by conducting sample surveys. However, it is not always possible to reduce non sampling error sufficiently to compensate for sampling error, as would be the case in a study involving automobile manufacturers In general, in marketing research projects dealing with consumers or households, a census is not

feasible and a sample should be taken. However, in many business- to-business or industrial marketing situations where the population size is small, a census is feasible as well as desirable.



CHECK YOUR PROGRESS

Q.1: Define Population and Sample

.....

.....

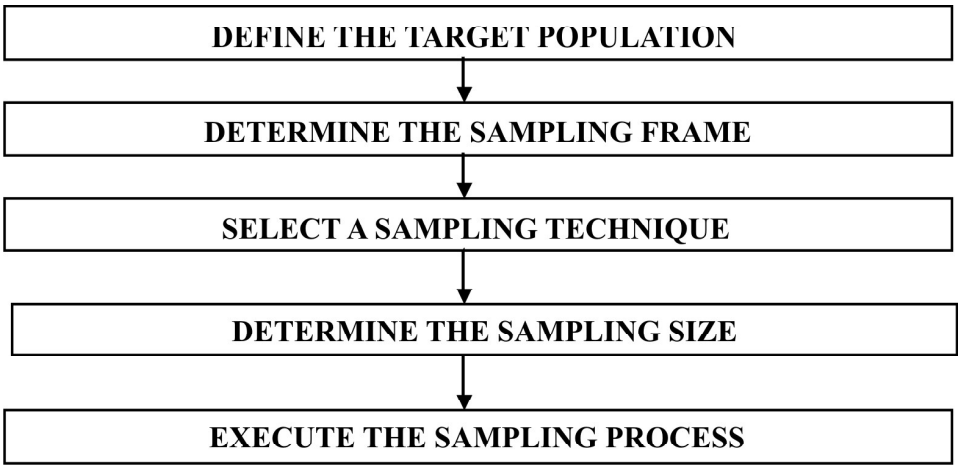
Q.2: What are Sampling Errors and Non Sampling Errors ?

.....

.....

7.5 SAMPLING DESIGN PROCESS

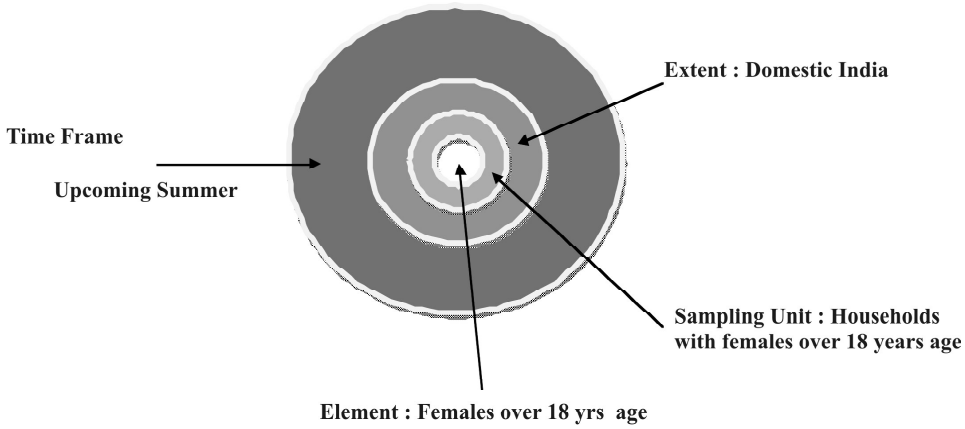
The sampling design process includes five steps, which are shown sequentially in the figure below. These steps are closely interrelated and relevant to all aspects of the marketing research project, from problem definition to the presentation of the results. Therefore, sample design decisions should be integrated with all other decisions in a research project.



1. Define the Target Population: Sampling design begins by specifying the target population. The **target population** is the collection of

elements or objects that possesses the information sought by the researcher and about which inferences are to be made. Inference involves projection or generalization of the sample results to the target population, as illustrated in the section entitled Classification of Sampling Techniques. The target population must be defined precisely. Imprecise definition of the target population will result in research that is ineffective at best and misleading at worst. Defining the target population involves translating the problem definition into a precise statement of who should and should not be included in the sample.

The target population should be defined in terms of elements***, sampling units, extent, and time***. An **element** is the object about which or from which the information is desired. In survey research, the element is usually the respondent. A **sampling unit** is an element, or a unit containing the element, that is available for selection at some stage of the sampling process, for example, a household. Suppose that Revlon wanted to assess consumer responses to a new line of lipsticks and wanted to sample females over 18 years of age. It may be possible to sample females over the age of 18 directly, in which case a sampling unit would be the same as an element. Alternatively, the sampling unit might be households. In the latter case, households would be sampled and all females over age 18 in each selected household would be interviewed or one selected from each sampled household, for example, by using the next birthday method. Here, the sampling unit and the population element are different. ***Extent*** refers to the geographical boundaries, and the ***time*** frame is the time period under consideration. Defining the target population is illustrated in the figure below.



2. Determine the Sampling Frame: 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. Examples of a sampling frame include the telephone book, an association directory listing the firms in an industry, a mailing list purchased from a commercial organization, a city directory, or a map. If a list cannot be compiled, then at least some directions for identifying the target population should be specified, such as random digit dialing procedures in telephone surveys. Often it is possible to compile or obtain a list of population elements, but the list may omit some elements of the population or include other elements that do not belong. Therefore, the use of a list will lead to sampling frame error. In some instances, the discrepancy between the population and the sampling frame is small enough to ignore. In most cases, however, the researcher should recognize and treat the sampling frame error.

3. Select a Sampling Technique: The most important decision about the choice of sampling technique is whether to use probability or non probability sampling. Given the importance of this decision, the issues involved are discussed in great detail later in this unit. If the sampling unit is different from the element, it is necessary to specify precisely how the elements within the sampling unit should be selected. In in-home personal interviews and telephone interviews, merely specifying the address or the telephone number may not be sufficient. For example, should the person answering the doorbell or the telephone be interviewed, or someone else in the household? Often, more than one person in a household may qualify. For example, both the male and female head of household may be eligible to participate in a study examining family leisure activities. When a probability sampling technique is being employed, a random selection must be made from all the eligible persons in each household. A simple procedure for random selection is the next birthday method.

4. Determine the Sample size: Sample size refers to the number of elements to be included in the study. Determining the sample size is complex and involves several qualitative and statistical considerations. Important qualitative factors that should be considered in determining the sample size include (1) the importance of the decision, (2) the nature of the research, (3) the nature of the analysis, (4) sample sizes used in similar studies, and (5) resource constraints.

In general, for more important decisions, more information is necessary and the information should be obtained more precisely. This calls for larger samples, but as the sample size increases, each unit of information is obtained at greater cost and less benefit. The nature of the research also has an impact on the sample size. For exploratory research designs, such as those using qualitative research, the sample size is typically small. For conclusive research, such as descriptive surveys, larger samples are required. If sophisticated analysis of the data using advanced statistical techniques is required, the sample size should be large. The same applies if the data are to be analyzed in great detail. Thus, a larger sample would be required if the data are being analyzed at the subgroup or segment level than if the analysis is limited to the aggregate or total sample. Sample size is influenced by the average size of samples in similar studies. Table 6.1 gives an idea of sample sizes used in different marketing research studies. These sample sizes have been determined based on experience and can serve as rough guidelines, particularly when non probability sampling techniques are used.

The sample size decision should be guided by a consideration of the resource constraints. In any marketing research project, money and time are limited. Other constraints include the availability of qualified personnel for data collection. Based on all the foregoing considerations, the sample size in a research project was determined to be 1,000. The statistical considerations involve the variability (variance) of the characteristic being measured and other considerations. Statistically, a sample of 400 (385 to be more precise) is sufficient to yield estimates of percentages that meet commonly acceptable statistical standards.

Table 7.1 Sample Sizes Used in Marketing Research Studies

Type of Study	Minimum Size	Typical Range
• Problem-identification research (e.g., market potential)	500	1,000–2,500
• Problem-solving research (e.g., pricing)	200	300–500
• Product tests	200	300–500
• Test marketing studies	200	300–500
• TV/radio/print advertising (per commercial or ad tested)	150	200–300
• Test-market audits	10 stores	10–20 stores
• Focus groups	2 groups	10–15 groups

Execute the Sampling Process

Execution of the sampling process requires a detailed specification of how the sampling design decisions with respect to the population, sampling frame, sampling unit, sampling technique, and sample size are to be implemented. If households are the sampling units, an operational definition of a household is needed. Procedures should be specified for vacant housing units and for call backs in case no one is at home. Detailed information must be provided for all sampling design decisions.



CHECK YOUR PROGRESS

Q.3: What is a Sampling Unit ?

.....

.....

Q.4: What is Sample Size.

.....

.....

7.6 NON PROBABILISTIC & PROBABILISTIC SAMPLING TECHNIQUES

Sampling techniques may be broadly classified as non probability and probability.

Non probability sampling relies on the personal judgment of the researcher rather than chance to select sample elements. The researcher can arbitrarily or consciously decide what elements to include in the sample. Non probability samples may yield good estimates of the population characteristics. However, they do not allow for objective evaluation of the sampling errors. Because there is no way of determining the probability of selecting any particular element for inclusion in the sample, the estimates obtained are not statistically projectable to the population; that is, one cannot derive

inferences about the population based on the sample results. Thus, if 35 percent of the people in a sample selected by non probability sampling express preference for a new service, we cannot conclude that 35 percent of the target population will have the same preference. Commonly used non probability sampling techniques include –

- **Convenience Sampling,**
- **Judgmental Sampling,**
- **Quota Sampling and**
- **Snowball Sampling.**

In **probability sampling**, sampling units are selected by chance. Because sample elements are selected by chance, it is possible to determine the accuracy of the sample estimates of the characteristics of interest, for example, the percentage expressing preference for a new service. This permits the researcher to make inferences or projections about the target population from which the sample was drawn. Thus, if 35 percent of the people in a sample selected by probability sampling express preference for a new service, we can conclude that it is likely that 35 percent of the target population will have the same preference. Probability sampling techniques consist of –

- **Simple Random Sampling,**
- **Systematic Sampling,**
- **Stratified Sampling and**
- **Cluster Sampling.**



CHECK YOUR PROGRESS

Q.5) What are the different types of Probability and Non Probability Sampling Techniques ?

.....

.....

7.6.1 Non-Probabilistic Sampling Techniques

1. Convenience Sampling: Convenience sampling attempts to obtain a sample of convenient elements. The selection of sampling units is left primarily to the interviewer. Often, respondents are selected because they happen to be in the right place at the right time. Examples of convenience sampling include (1) use of students, church groups, and members of social organizations; (2) mall-intercept interviews without qualifying the respondents; (3) department stores using charge account lists; and (4) "people on the street" interviews. Convenience sampling is the least expensive and least time-consuming of all sampling techniques. The sampling units are accessible, easy to measure, and cooperative. In spite of these advantages, this form of sampling has serious limitations. Many potential sources of selection bias are present, including respondent self-selection (that is, respondents can influence their selection or non selection). Convenience samples are not representative of any definable population. Hence, it is not meaningful to generalize the sample results to any population from a convenience sample, and convenience samples are not appropriate for marketing research projects involving population inferences. Convenience samples are not recommended for descriptive or causal research, but they can be used in exploratory research for generating ideas, insights, or hypotheses. Convenience samples can be used for focus groups, pretesting questionnaires, or pilot studies. Even in these cases, caution should be exercised in interpreting the results.

2. Judgmental Sampling: Judgmental sampling is a form of convenience sampling in which the population elements are selected based on the judgment of the researcher. The researcher, exercising judgment or expertise, chooses the elements to be included in the sample because he or she believes that they are representative of the population of interest or are otherwise appropriate. Common examples of judgmental sampling include (1) test markets selected to determine the potential of a new product, (2) purchase engineers selected in industrial marketing research because they are considered

to be representative of their respective companies, and (3) bellwether precincts selected in voting behavior research.

3. Quota Sampling: Quota sampling may be viewed as two-stage judgmental sampling. The first stage consists of developing control categories, or quotas, of population elements. To develop these quotas, the researcher lists relevant control characteristics and determines the distribution of these characteristics in the target population. The relevant control characteristics, which may include gender, age, and race, are identified on the basis of judgment. Often, the quotas are assigned so that the proportion of the sample elements possessing the control characteristics is the same as the proportion of population elements with these characteristics. In other words, the quotas ensure that the composition of the sample is the same as the composition of the population with respect to the characteristics of interest. In the second stage, sample elements are selected based on convenience or judgment. Once the quotas have been assigned, there is considerable freedom in selecting the elements to be included in the sample. The only requirement is that the elements selected fit the control characteristics.

Even if the sample composition mirrors that of the population with respect to the control characteristics, there is no assurance that the sample is representative. If a characteristic that is relevant to the problem is overlooked, the quota sample will not be representative. Relevant control characteristics are often omitted because practical difficulties are associated with trying to include many control characteristics. Because the elements within each quota are selected based on convenience or judgment, many sources of selection bias are potentially present. The interviewers may choose to go to selected areas where eligible respondents are more likely to be found. They may avoid people who look unfriendly or are not well dressed, or those who live in undesirable locations. Quota sampling does not permit assessment of sampling error. Quota sampling attempts to obtain representative samples at a relatively low cost. Its advantages are the lower costs and greater convenience to the interviewers in selecting elements for each quota. Recently, tighter controls have

been imposed on interviewers and interviewing procedures that tend to reduce selection bias, and guidelines have been suggested for improving the quality of mall-intercept quota samples. Under certain conditions, quota sampling obtains results close to those for conventional probability sampling.

4. Snowball Sampling : In snowball sampling, an initial group of respondents is selected, usually at random. After being interviewed, these respondents are asked to identify others who belong to the target population of interest. Subsequent respondents are selected based on the referrals. This process may be carried out in waves by obtaining referrals from referrals, thus leading to a snowball effect. Even though probability sampling is used to select the initial respondents, the final sample is a non probability sample. The referrals will have demographic and psychographic characteristics that are more similar to the persons referring them than would occur by chance.

A major objective of snowball sampling is to estimate characteristics that are rare in the population. Examples include users of particular government or social services, such as food stamps, whose names cannot be revealed; special census groups, like widowed males under age 35; and members of a scattered minority population. Snowball sampling is used in industrial buyer–seller research to identify buyer–seller pairs. The major advantage of snowball sampling is that it substantially increases the likelihood of locating the desired characteristic in the population. It also results in relatively low sampling variance and costs.

7.6.2 Probabilistic Sampling Technique

Probability sampling techniques vary in terms of sampling efficiency. Sampling - efficiency reflects a trade-off between sampling cost and precision. Precision refers to the level of uncertainty about the characteristic being measured. Precision is inversely related to sampling errors but positively related to cost. The greater the precision, the lower the sampling errors and the greater the cost, and most

studies require a trade-off. The researcher should strive for the most efficient sampling design, subject to the budget allocated. The efficiency of a probability sampling technique may be assessed by comparing it to that of simple random sampling.

1. Simple Random Sampling: In simple random sampling (SRS), each element in the population has a known and equal probability of selection, and every element is selected independently of every other element. The sample is drawn by a random procedure from a sampling frame. This method is equivalent to a lottery system in which names are placed in a container, the container is shaken, and the names of the winners are then drawn in an unbiased manner.

To draw a simple random sample, the researcher first compiles a sampling frame in which each element is assigned a unique identification number. Then random numbers are generated to determine which elements to include in the sample. The random numbers may be generated with a computer routine or a table. This was illustrated in a Case where a computer program was used to randomly generate household telephone numbers. SRS has many desirable features. It is easy to understand. The sample results may be projected to the target population. Most approaches to statistical inference assume that the data have been collected by simple random sampling. However, SRS suffers from at least four significant limitations. **First,** it is often difficult to construct a sampling frame that will permit a simple random sample to be drawn. **Second,** SRS can result in samples that are very large or spread over large geographic areas, thus increasing the time and cost of data collection. **Third,** SRS often results in lower precision than other probability sampling techniques. In fact, the efficiency of other probability sampling techniques is often assessed by comparing it to that of simple random sampling (SRS). **Fourth,** SRS may or may not result in a representative sample. Although samples drawn will represent the population well on average, a given simple random sample may grossly misrepresent the target population. This is more likely if the size of the sample is small. For these reasons, SRS is not widely

used in marketing research. Procedures such as systematic sampling are more popular.

2. Systematic Sampling: In systematic sampling, the sample is chosen by selecting a random starting point and then picking every *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, there are 100,000 elements in the population, and a sample of 1,000 is desired. In this case, the sampling interval, *i*, is 100. A random number between 1 and 100 is selected. If this number is 23, for example, the sample consists of elements 23, 123, 223, 323, 423, 523, and so on.

Systematic sampling is different from SRS in that only the initial element is selected randomly and the remaining elements are selected systematically, that is, by adding *i* to the previous element number. In SRS, all the elements are selected randomly.

For systematic sampling, the researcher assumes that the population elements are ordered in some respect. When the ordering of the elements is related to the characteristic of interest, systematic sampling increases the representativeness of the sample. If firms in an industry are arranged in increasing order of annual sales, a systematic sample will include some small and some large firms. A simple random sample may be unrepresentative because it may contain, for example, only small firms or a disproportionate number of small firms. If the ordering of the elements produces a cyclical pattern, systematic sampling may decrease the representativeness of the sample. To illustrate, consider the use of systematic sampling to generate a sample of monthly department store sales from a sampling frame containing monthly sales for the last sixty years. If a sampling interval of 12 is chosen, the resulting sample would not reflect the month-to-month variation in sales. Let's say that a random number from 1 to 12 is selected, and let's say that 3 is the number selected, representing the month of March. Thus, the sample will consist of 3, 15 (3 + 12), 27 (3 + [2 × 12]), and the month of March will

be selected for each of the sixty years. In this case, the monthly variation in sales for each year would not be reflected in the sample.

Systematic sampling is less costly and easier than SRS because random selection is done only once. The random numbers do not have to be matched with individual elements, as in SRS. Another relative advantage is that systematic sampling can even be used without knowledge of the composition (elements) of the sampling frame. For example, every *i*th person leaving a department store or mall can be intercepted. For these reasons, systematic sampling is often employed in consumer mail, telephone, and mall-intercept interviews

Stratified Sampling: Stratified sampling is a two-step process in which the population is first partitioned into subpopulations, or strata. The strata should be mutually exclusive and collectively exhaustive in that every population element should be assigned to one and only one stratum and no population elements should be omitted. Next, elements are selected from each stratum by a random procedure, usually simple random sampling (SRS). Technically, only SRS should be employed in selecting the elements from each stratum. In practice, sometimes systematic sampling and other probability sampling procedures are employed. Stratified sampling differs from quota sampling in that the sample elements are selected probabilistically rather than based on convenience or judgment. A major objective of stratified sampling is to increase precision without increasing cost.

The variables used to partition the population into strata are referred to as stratification variables. The criteria for the selection of these variables consist of homogeneity (similarity), heterogeneity (dissimilarity), relatedness, and cost. The elements within a stratum should be as homogeneous (similar) as possible, but the elements in different strata should be as heterogeneous (different) as possible. The stratification variables should also be closely related to the characteristic of interest. The more closely these criteria are met, the greater the effectiveness in controlling extraneous sampling variation.

Finally, the variables should decrease the cost of the stratification process by being easy to measure and apply.

Variables commonly used for stratification include demographic characteristics (as illustrated in the example for quota sampling), type of customer (credit card versus non–credit card), size of firm, or type of industry. It is possible to use more than one variable for stratification, although more than two are seldom used because of pragmatic and cost considerations.

3. Cluster Sampling: In cluster sampling, the target population is first divided into mutually exclusive and collectively exhaustive subpopulations, or clusters. Then a random sample of clusters is selected, based on a probability sampling technique such as simple random sampling. For each selected cluster, either all the elements are included in the sample or a sample of elements is drawn probabilistically. If all the elements in each selected cluster are included in the sample, the procedure is called one-stage cluster sampling. If a sample of elements is drawn probabilistically from each selected cluster, the procedure is two-stage cluster sampling.

The key distinction between cluster sampling and stratified sampling is that, in cluster sampling, only a sample of subpopulations (clusters) is chosen, whereas in stratified sampling, all the subpopulations (strata) are selected for further sampling. The objectives of the two methods are also different. The objective of cluster sampling is to increase sampling efficiency by decreasing costs. The objective of stratified sampling is to increase precision. With respect to homogeneity and heterogeneity, the criteria for forming clusters are just the opposite of that for strata. Elements within a cluster should be as heterogeneous (different) as possible, but clusters themselves should be as homogeneous (similar) as possible. Ideally, each cluster should be a small-scale representation of the population. In cluster sampling, a sampling frame is needed only for those clusters selected for the sample. A common form of cluster sampling is area sampling, in which the clusters consist of geographic areas, such as counties, housing tracts, or blocks.



CHECK YOUR PROGRESS

Q.6: Write note on Convenience Sampling and Snowball Sampling

.....

.....

Q.7: Discuss about Simple Random Sampling and Cluster Sampling

.....

.....

7.7 DETERMINATION OF SAMPLE SIZE

Sample Size determination is computed using three inputs:

- The estimate of the population standard deviation (often obtained from earlier studies).
- The acceptable level of sampling error
- The desired confidence level

Generally, research practitioners utilize the following sequence and inputs in computing sample size:

1. Survey respondents will split 50/50 in response to dichotomous (e.g. yes/no) questions.
2. The desired level of confidence will be 95%, or 1.96 standard deviations from the mean or .05 possible

Py = Proportion responding “yes”

Pn = Proportion responding “no”

Standard error is the acceptable amount of error/confidence interval. Let us consider it as .05/1.96 (about 2 standard deviations), or .0255102.

The standard formula for computing the sample size is:

$$\frac{(Py)\,(Pn)}{\text{Std Error}^2}$$

So, when the respective values are input, we end up with .25/.0006507 or 384 respondents. This is why a survey sample size of 400 is often recommended.

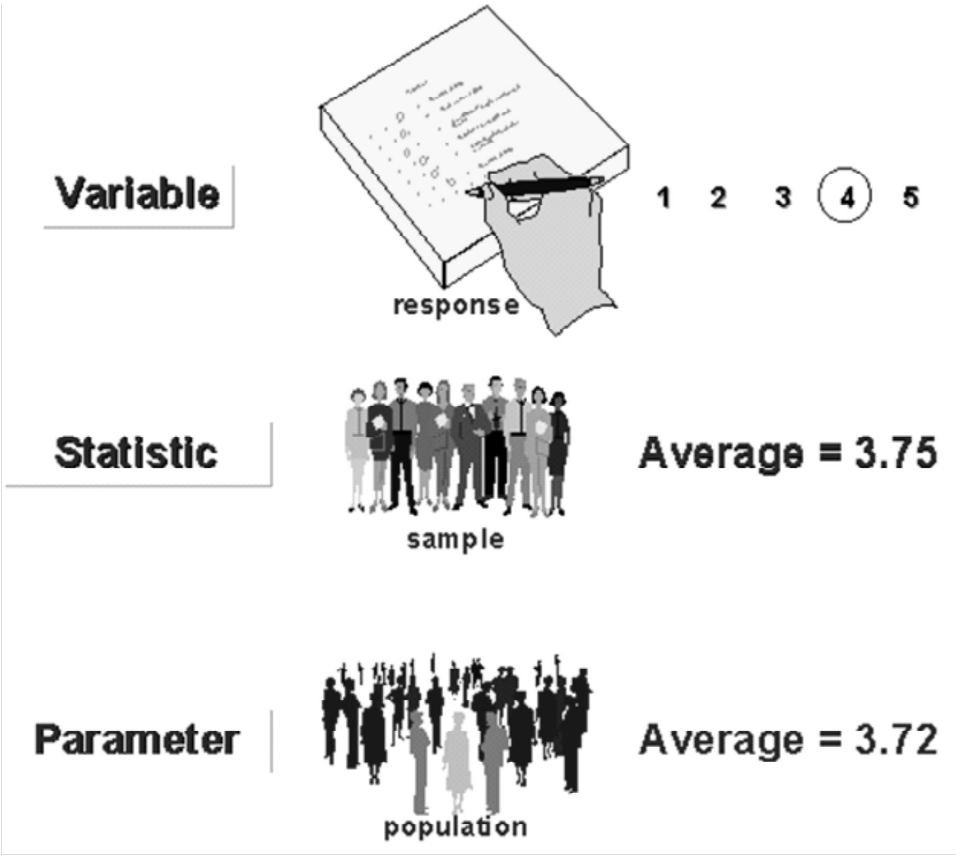
Sample size is important in avoiding Type I or Type II errors.
Type I errors are made by stating that there is a difference between two groups within a population on a given measurement, when in fact there is no difference. Accommodating this potential outcome is where most sample size calculations stop. Often, practitioners simply ignore the possibility of making a Type II error.

The sample size typically needed to address Type I errors is 384.
Type II errors are made by stating that there is no difference between two groups within a population on a given measurement, when in fact there is a difference. While important, many researchers ignore statistical power calculations. In the "real world" tables and canned statistical tools are utilized to determine survey power, due to the complexity of the formulas.

The sample size typically needed to address Type II errors is 1,236.
Confidence level suggests that other samples drawn from the same population will have similar values X% of the time. For most marketing research exercises, confidence levels are set at 95%.
Confidence interval includes the possible end point values for the entire population. The confidence interval allows for a computed amount of variation from the mean value based on the precision/cost value trade-off.

Sampling Error :Any characteristic of the population related measure is called a parameter. It is a fixed value as long as the population does not change.

The picture shows population. Say it includes 1000 people and the average of their response is 3.72. It is known as population mean –a parameter.



The picture also shows a sample. Let us say, it consists of 100 people. The sample mean is 3.75 and it is called a statistic. The difference between the statistic and the parameter is called sampling error. Each sample will generate a different amount of sampling error. In this case it is 3.75-3.72 = 0.03.

Sampling Distribution of Mean: The distribution of responses of sample elements form a sampling distribution with a mean and variations. To understand, let us to do an exercise. Imagine the population of 5 elements say [1, 2, 3, 4, 5]. The mean value of the population = [1+2+3+4+5]/5 =3

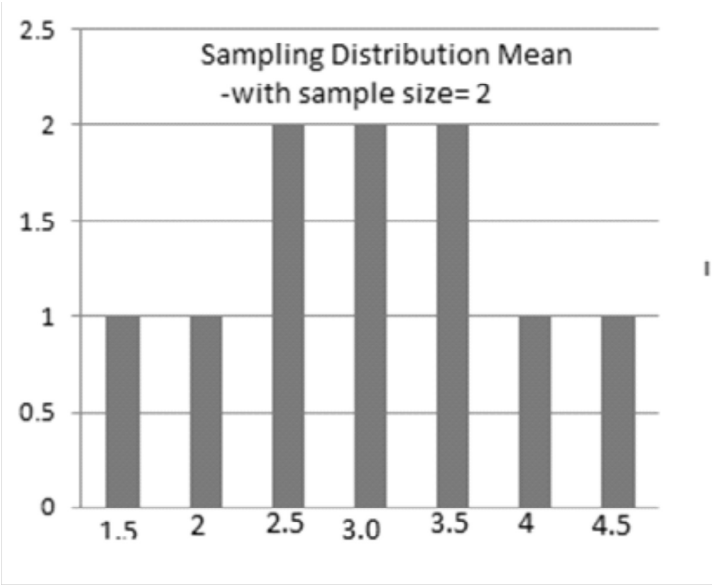
Let us assume that you have taken sample sizes of 2, 3 and 4.

Sample Size -2 units

For the sample size of 2, you know that a maximum of 10 possible unique samples can be selected [5C2 = 5 *4 /1 * 2 = 10]. If you look at a glance of this distribution of sampling means, it is varying from 1.5 to 4.5.

SL.NO	SAMPLE ELEMENTS	MEAN OF THE SAMPLE
1	1,2	1.5
2	1,3	2
3	1.4	2.5
4	1,5	3
5	2,3	2.5
6	2,4	3
7	2,5	3.5
8	3,4	3.5
9	3,5	4
10	4,5	4.5

Distribution of Sampling Means for a sample size 2

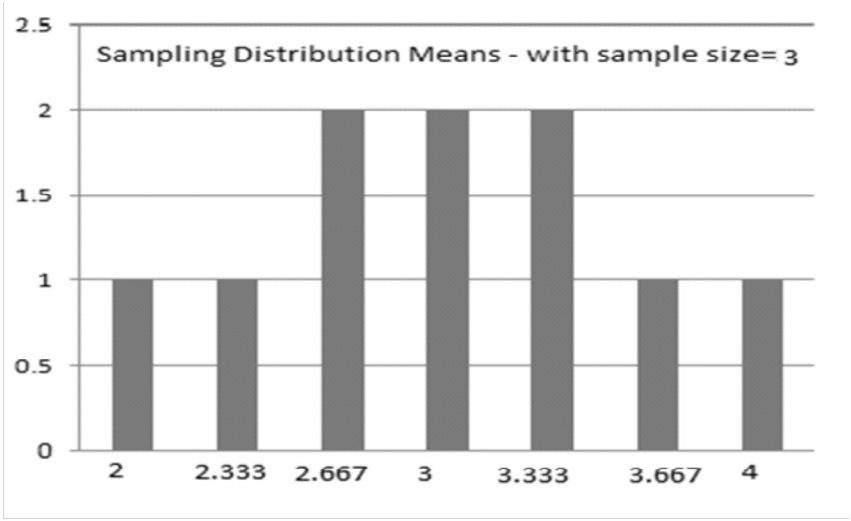


Sample Size of 3

A maximum of 10 possible unique samples can be selected [5C3 = (5 * 4 * 3) / (1*2*3) = 10]. If you look at a glance of this distribution of sampling means, it is varying from 2.0 to 4.0

SL.NO	SAMPLE ELEMENTS	MEAN OF THE SAMPLE
1	1,2,3	2
2	1,2,4	2.3333
3	1.2.5	2.6666
4	1,,3,4	2.666
5	1,3,5	3
6	1,4,5	3.3333
7	2,3,4	3
8	2,3,5	3.3333
9	2,4,5	3.666
10	3,4,5	4

Distribution of Sampling Means for a sample size 3

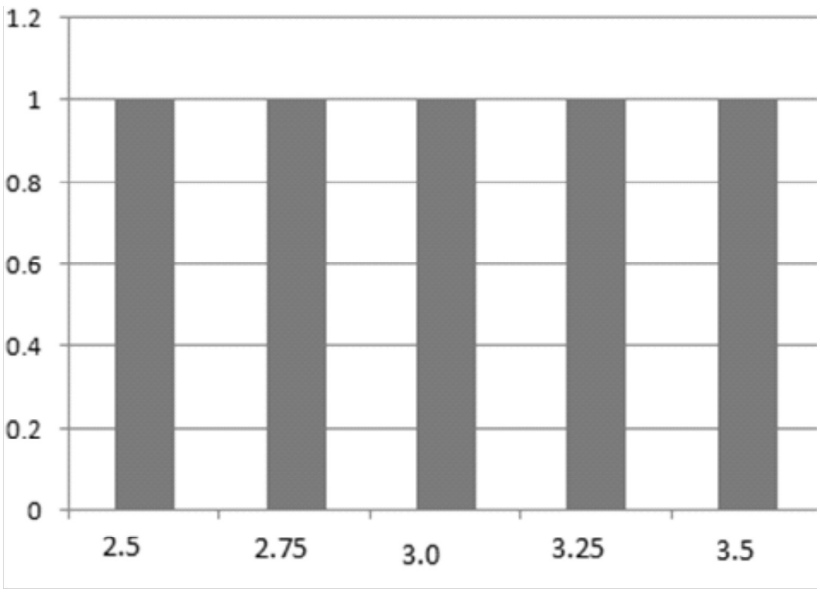


Sample Size of 4

A maximum of 5 possible unique samples can be selected [5C4 = (5 * 4 * 3 * 2) / (1*2*3*4) = 5]. If you look at a glance of this distribution of sampling means, it is varying from 2.5 to 3.5.

SL.NO	SAMPLE ELEMENTS	MEAN OF THE SAMPLE
1	1,2,3,4	2.5
2	1,2,4,5	3
3	1,3,4,5	3.25
4	1,2,3,5	2.75
5	2,3,4,5	3.5

Distribution of Sampling Means for a sample size 3



In each case the average of means is equal to 3. Therefore we can say, for a finite population, different sample may or may not have mean equal to population. But the mean of the distribution will be equal to population mean. The standard deviation of each of these samples represents the distribution of potential sampling error, which in statistical terminology is known as standard error.

Even though all three samples came from the same population, you would not expect to get the exact same statistic from each. They would differ marginally just due to the random “luck of the draw” or to the natural fluctuations or vagaries of drawing a sample. But you would expect that all

three samples would yield a similar statistical estimate because they were drawn from the same population.

Now, if we extend the same logic to an infinite number of samples from a population, which consist of 1000s of members. Even for a sample size of 100, [1000C100 = you can calculate...] countable samples [of course in millions] can be drawn, which is theoretically possible to do and if we compute the average for each one.

If you plotted them on a histogram or bar graph you should find that most of them converge on the same central value and that you get fewer and fewer samples that have averages farther away up or down from that central value. In other words, the bar graph would be well described by the bell curve shape that is an indication of a “normal” distribution in statistics. The distribution of an infinite number of samples of the same size as the sample in your study is known as the sampling distribution of means.



7.8 LET US SUM UP

Sampling is preferred when the cost of sampling error is low, and the cost of non sampling error is high. The opposite set of conditions favor the use of a census. Sampling design begins by defining the target population in terms of elements, sampling units, extent, and time. Then the sampling frame should be determined. A sampling frame is a representation of the elements of the target population. It consists of a list or directions for identifying the target population. At this stage, it is important to recognize any sampling frame errors that may exist. The next step involves selecting a sampling technique and determining the sample size. In addition to quantitative analysis, several qualitative considerations should be taken into account in determining the sample size. Finally, execution of the sampling process requires detailed specifications for each step in the sampling process.

Sampling techniques may be classified as non probability and probability techniques. Non probability sampling techniques rely on the researcher’s judgment. Consequently, they do not permit an objective

evaluation of the precision of the sample results, and the estimates obtained are not statistically projectable to the population. The commonly used non probability sampling techniques include convenience sampling, judgmental sampling, quota sampling, and snowball sampling.

In probability sampling techniques, sampling units are selected by chance. Each sampling unit has a nonzero chance of being selected. It is also possible to determine the precision of the sample estimates and to make projections to the target population. Probability sampling techniques include simple random sampling, systematic sampling, stratified sampling, and cluster sampling. The choice between probability and non probability sampling should be based on the nature of the research, relative magnitude of sampling and non sampling errors, and statistical and operational considerations.



7.9 FURTHER READING

Malhotra, N.K. and Dash, S. (2011). *Marketing research: An Applied Orientation* (6th edition), New Delhi: Pearson Education.



7.10 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q1: A **population** is the total of all the elements that share some common set of characteristics. A **sample** is a subgroup of the population selected for the marketing research project.

Ans to Q2: **Sampling errors** are errors resulting from the particular sample selected being an imperfect representation of the population of interest. **Non Sampling errors,** on the other hand, can be attributed to sources other than sampling.

Ans to Q3: A **sampling unit** is an element, or a unit containing the element, that is available for selection at some stage of the sampling process.

Ans to Q4: Sample size refers to the number of elements to be included in the study.Important qualitative factors that should be considered in determining the sample size include (1) the importance of the decision, (2) the nature of the research, (3) the nature of the analysis, (4) sample sizes used in similar studies, and (5) resource constraints.

Ans to Q5: Probability sampling techniques consist of :Simple Random Sampling, Systematic Sampling, Stratified Sampling, and Cluster Sampling.

Non probability sampling techniques include:Convenience Sampling, Judgmental Sampling, Quota Sampling and Snowball Sampling.

Ans to Q6. Convenience sampling attempts to obtain a sample of convenient elements. The selection of sampling units is left primarily to the interviewer. Often, respondents are selected because they happen to be in the right place at the right time.

In snowball sampling, an initial group of respondents is selected, usually at random. After being interviewed, these respondents are asked to identify others who belong to the target population of interest.

Ans to Q7. In simple random sampling (SRS), each element in the population has a known and equal probability of selection, and every element is selected independently of every other element.

In **cluster sampling**, the target population is first divided into mutually exclusive and collectively exhaustive subpopulations, or clusters.



7.11 MODEL QUESTIONS

1) What is the major difference between a sample and a census?
2) Under what conditions would a sample be preferable to a census? A census preferable to a sample?
3) Describe the sampling design process.
4) How should the target population be defined?
5) What is a sampling unit? How is it different from the population element?
6) What qualitative factors should be considered in determining the sample size?
7) What are the disadvantages of non probability sampling techniques?
8) What is the least expensive and least time-consuming of all sampling techniques? What are the major limitations of this technique?
9) What are the advantages and disadvantages of judgmental sampling?
10) What is the relationship between quota sampling and judgmental sampling?
11) What are the distinguishing features of simple random sampling?
12) Describe the procedure for selecting a systematic random sample.
13) Describe stratified sampling. What are the criteria for the selection of stratification variables?
14) What is the major advantage of snowball sampling?
15) Discuss the method of determining the sample size using means and proportions.

*****