Prepared By:
Dr. Gaurav Sehgal
Associate Professor
School of Business Studies
Department of HRM & OB
Central University of Jammu, Jammu, J&K State

UNIT – III
Topics Covered

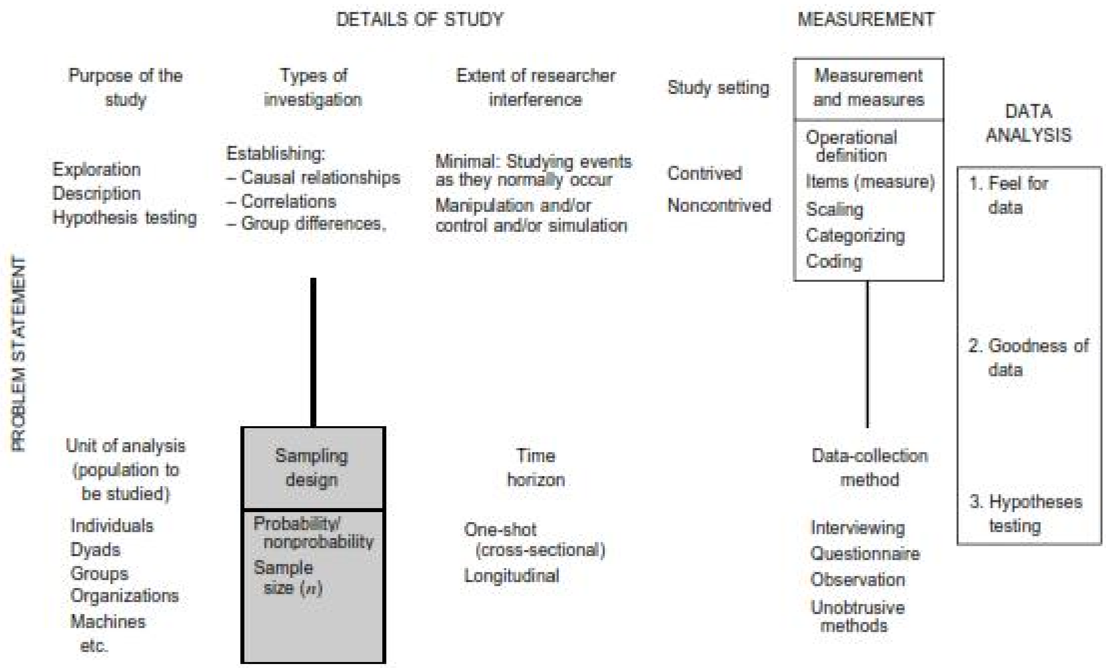
- ❖ Stages of Questionnaire Development
- ❖ Sample Unit
- ❖ Sampling Techniques – Probability & Non-probability
- ❖ Sample Size and it's calculations
- ❖ Procedure to select a sample
- ❖ Sampling and Non-sampling Error
- ❖ Methods of Data Collection – Primary and Secondary data
- ❖ Data collection through Questionnaire and Schedules

Prepared By:
Dr. Gaurav Sehgal
Associate Professor
School of Business Studies
Department of HRM & OB
Central University of Jammu, Jammu, J&K State

3.1 Introduction

Surveys are useful and powerful in finding answers to research questions through data collection and subsequent analyses, but they can do more harm than good if the population is not correctly targeted. That is, if data are not collected from the people, events or objects that can provide the correct answers to solve the problem, the survey would be in vain.

The process of selecting the right individuals, objects or events for study is known as *Sampling* (shaded portion in the figure given below).



3.2 Important terms

(i) ***Population:*** Population refers to the entire group of people, events or things of interest that the researcher wishes to investigate. For instance, if the CEO of a computer firm wants to know the kinds of advertising strategies adopted by computer firms in a particular Market-X, then all computer firms operating / situated there will be the population. If a banker is interested in investigating the savings habits of blue-collar workers in the mining industry in India, then all blue-collar workers in that industry throughout the country will form the population. If an organizational consultant is interested in studying the effects of a 4 day work-week on the white-collar workers in a telephone company in Hyderabad, then all white-collar workers in that company will make up the population. If regulators want to know how patients in nursing homes run by Apollo Hospitals are cared for, then all the patients in all the nursing homes run by them will form the population. If, however, the regulators are interested only in one particular nursing home in Delhi run by Apollo Group, then only the patients in that specific nursing home will form the population.

(ii) ***Element:*** An element is a single member of the population. If 1,000 blue-collar workers in a particular organization happen to be the population of interest to a researcher, each blue-collar worker therein is an element. If 500 pieces of machinery are to be approved after inspecting a few, there would be 500 elements in this population. Incidentally, the census is a count of all elements in the human population.

(iii) ***Population Frame:*** The population frame is a listing of all the elements in the population from which the sample is drawn. The payroll of an organization would serve as the population frame if its members are to be studied. Likewise, the university registry containing a listing of all students, faculty, administrators and support staff in the university during a particular academic year or semester could serve as the population frame for a study of the university population. A roster of class students could be the population frame for the study of students in a class. The telephone directory is also frequently used as a population frame for some types of studies, even though it has an inherent bias in as much as some numbers are unlisted and certain others may have become obsolete.

Although the population frame is useful in providing a listing of each element in the population, it may not always be a current, updated document. For instance, the names of members who have recently left the organization or dropped out of the university, as well as members who have only recently joined the organization or the university may not appear in the organization‘s payroll or the university registers on a given day. The most recently installed or disconnected telephones will not, likewise, be included in the current telephone directory. Hence, though the population frame may be available in many cases, it may not always be entirely correct or complete. However, the researcher might recognize this problem and not be too concerned about it, because a few additions and deletions in the telephone directory might not make any significant difference to the study. Even if she is concerned about it, and spends time and effort trying to obtain an updated population frame, there is no guarantee that the new population frame has an accurate listing of *all* the elements either.

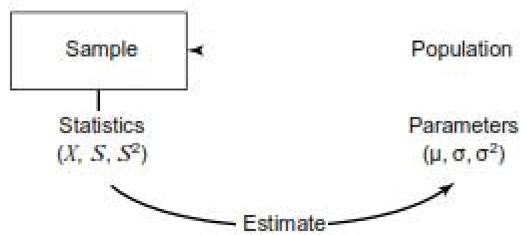
(iv) ***Sample:*** A sample is a subset of the population. It comprises some members selected from it. In other words, some, but not all, elements of the population would form the sample. If 200 members are drawn from a population of 1,000 blue-collar workers, these 200 members form the sample for the study. That is, from a study of these 200 members, the researcher would draw conclusions about the entire population of the 1,000 blue-collar workers. Likewise, if there are 145 in-patients in a hospital and 40 of them are to be surveyed by the hospital administrator to assess their level of satisfaction with the treatment received, then these 40 members will be the sample. A sample is thus a subgroup or subset of the population. By studying the sample, the researcher should be able to draw conclusions that would be generalizable to the population of interest.

Prepared By:
Dr. Gaurav Sehgal
Associate Professor
School of Business Studies
Department of HRM & OB
Central University of Jammu, Jammu, J&K State

3.3 Sampling and Reasons for Sampling

(i) Sampling: Sampling is the *process* of selecting a sufficient number of elements from the population, so that a study of the sample and an understanding of its properties or characteristics would make it possible for us to generalize such properties or characteristics to the population elements. The characteristics of the population such as μ (the population mean), σ (the population standard deviation) and σ^2 (the population variance) are referred to as its *parameters*. The central tendencies, the dispersions and other statistics in the sample of interest to the research are treated as approximations of the central tendencies, dispersions and other parameters of the population. As such, all conclusions drawn about the sample under study are generalized to the population. In other words, the sample statistics $\overline{X}$ (the sample mean), *S* (standard deviation), and S^2 (the variation in the sample), are used as estimates of the population parameters μ, S and S^2. Figure below shows the relationship between the sample and the population.

The Relationship between Sample and Population.



(ii) Reasons for Sampling: The reasons for using a sample, rather than collecting data from the entire population, are self-evident. In research investigations involving several hundreds and even thousands of elements, it would be practically impossible to collect data from, or test, or examine every element. Even if it were possible, it would be prohibitive in terms of time, cost and other human resources. Study of a sample rather than the entire population is also sometimes likely to produce more reliable results. This is mostly because fatigue is reduced and fewer errors will therefore result in collecting data, especially when a large number of elements is involved. In a few cases, it would also be impossible to use the entire population to gain knowledge about, or test something. Consider, for instance, the case of electric bulbs. In testing the life of a batch of bulbs, if we were to burn every bulb produced, there would be none left to sell! This is known as destructive sampling.

(iii) Representativeness of Samples: The need for choosing the right sample for a research investigation cannot be overemphasized. We know that rarely will the sample be the exact replica of the population from which it is drawn. For instance, very few sample means ($\overline{X}$) likely to be exactly equal to the population means (μ). Nor is the standard deviation of the sample (S) likely to be the same as the standard deviation of the population (σ). However, if we choose the sample in a scientific way, we can be reasonably sure that the sample statistic (e.g., $\overline{X}$, S or S^2) is fairly close to the population parameter (i.e., μ, σ or σ^2). To put it differently, it is possible to choose the sample in such a way that it is representative of the population. There is always the slight probability, however, that sample values might fall outside the population parameters.

3.4 Sampling Designs

There are two major types of sampling designs: Probability and Non-probability sampling.

In probability sampling, the elements in the population have some known chance or probability of being selected as sample subjects.

In non-probability sampling, the elements do not have a known or predetermined chance of being selected as subjects.

Probability sampling designs are used when the representativeness of the sample is of importance in the interests of wider generalizability. When time or other factors, rather than generalizability, become critical, non-probability sampling is generally used.

Each of these two major designs has different sampling strategies. Depending on the extent of generalizability desired, the demands of time and other resources, and the purpose of the study, different types of probability and non-probability sampling designs are chosen.

We shall now discuss these in detail.

3.5 Probability Sampling

When elements in the population have a known chance of being chosen as subjects in the sample, we resort to a probability sampling design. Probability sampling can be either unrestricted (simple random sampling) or restricted (complex probability sampling) in nature. In other words, there are two basic probability sampling plans: the unrestricted or simple random sampling, and the restricted or complex probability sampling plans. In the simple random sampling design, every element in the population has a known and equal chance of being selected as a subject. The complex probability plan consists of five different sampling designs. Of these five, the cluster sampling design is probably the least expensive as well as the least dependable, but is used when no list of the population elements is available. The stratified random sampling design is probably the most efficient, in the sense that for the same number of sample subjects, it offers precise and detailed information. The systematic sampling design has the built-in hazard of possible systematic bias. Area sampling is a popular form of cluster sampling, and double sampling is resorted to when information in addition to that already obtained by using a primary sample has to be collected using a subgroup of the sample.

(i) ***Unrestricted or Simple Random Sampling:*** In the unrestricted probability sampling design, more commonly known as simple random sampling, every element in the population has a *known and equal* chance of being selected as a subject. Let us say there are 1,000 elements in the population, and we need a sample of 100. Suppose we were to drop pieces of paper in a hat, each bearing the name of one of the elements, and draw 100 of those from the hat with our eyes closed. We know that the first piece drawn will have a 1/1,000 chance of being drawn, the next one a 1/999 chance of being drawn, and so on. In other words, we know that the probability of any one of them being chosen is 1 in the number of the population, and we also know that each single element in the hat has the same or equal probability of being chosen. We certainly know that computers can generate random numbers and one does not have to go through the tedious process of pulling out names from a hat! When we thus draw the elements from the population, it is most likely that the distribution patterns of the characteristics we are interested in investigating in the population are also likewise distributed in the subjects we draw for our sample. This sampling design, known as *simple random sampling*, has the least bias and offers the most generalizability. However, this sampling process could become cumbersome and expensive; in addition an entirely updated listing of the population may not always be available. For these and other reasons, other probability sampling designs are often chosen instead.

(ii) ***Restricted or Complex Probability Sampling:*** As an alternative to the simple random sampling design, several complex probability sampling (restricted probability) designs can be used. These probability sampling procedures offer a viable, and sometimes more efficient alternative to the unrestricted design we just discussed. Efficiency is improved in that more information can be obtained for a given sample size using some of the complex probability sampling procedures than the simple random sampling design.

The five most common complex probability sampling designs are - Systematic Sampling, Stratified Random Sampling, Cluster Sampling, Area Sampling and Double Sampling.

(a) Systematic Sampling

The systematic sampling design involves drawing every *n*th element in the population starting with a randomly chosen element between 1 and *n*. The procedure is as illustrated below.

Illustration-1: If we want a sample of 35 households from a total population of 260 houses in a particular locality, then we could sample every seventh house starting from a random number from 1 to 7. Let us say that the random number is 7, then houses numbered 7, 14, 21, 28, and so on, would be sampled until the 35 houses are selected. The one problem to be borne in mind in the systematic sampling design is the probability of a systematic bias creeping into the sample. In the above example, for instance, let us say that every seventh house happens to be a corner house. If the focus of the research study conducted by the construction industry is to control - noise pollution‖ experienced by residents through the use of appropriate filtering materials, then the residents of corner houses may not be exposed to as much noise as the houses that are in between. Information on noise levels gathered from corner house dwellers might therefore bias the researcher‘s data. The likelihood of drawing incorrect conclusions from such data is thus high. In view of the scope for such systematic bias, the researcher must consider the plans carefully and make sure that the systematic sampling design is appropriate for the study, before deciding on it.

For market surveys, consumer attitude surveys, and the like, the systematic sampling design is often used, and the telephone directory frequently serves as the population frame for this sampling design.

(b) Stratified Random Sampling

While sampling helps to estimate population parameters, there may be identifiable subgroups of elements within the population that may be expected to have different parameters on a variable of interest to the researcher. For example, to the Human Resource Management Director might be interested in assessing the extent of training that the employees in the system feel they need, the entire organization will form the population for study. But the extent, quality and intensity of training desired by middle-level managers, lower-level managers, first-line supervisors, computer analysts, clerical workers, and so on will be different for each group. Knowledge of the kinds of differences in needs that exist for the different groups will help the director to develop useful and meaningful training programs for each group in the organization. Data will therefore have to be collected in a manner that would help the assessment of needs at each subgroup level in the population. The unit of analysis then would be at the group level and the stratified random sampling process will come in handy.

Stratified random sampling, as its name implies, involves a process of stratification or segregation, followed by random selection of subjects from each stratum. The population is first divided into mutually exclusive groups that are relevant, appropriate, and meaningful in the context of the study. For instance, if the president of a company is concerned about low motivational levels or high absenteeism rates among the employees, it makes sense to stratify the population of organizational members according to their job levels. When the data are collected and the analysis is done, we may find that contrary to expectations, it is the middle-level managers that are not motivated. This information will help the president to focus on action at the right level and devise better methods to motivate this group. Tracing the differences in the parameters of the subgroups within a population would not have been possible without the stratified random sampling procedure. If either the simple random sampling or the systematic sampling procedure were used in a case like this, then the high motivation at some job levels and the low motivation at other levels would have canceled each other out, thus masking the real problems that exist at a particular level or levels.

Stratification also helps when research questions such as the following are to be answered: (1) Are the machinists more accident prone than clerical workers?; (2) Are Hispanics more loyal to the organization than Native Americans?

Stratifying customers on the basis of life stages, income levels, and the like to study buying patterns and stratifying companies according to size, industry, profits, and so forth to study stock market reactions are all common examples of the use of stratification as a sampling design technique.

Stratification is an efficient research sampling design; that is, it provides more information with a given sample size. Stratification should follow the lines appropriate to the research question. If we study consumer preferences for a product, stratification of the population could be by geographical areas, market segments, consumers' age, consumers' gender or various combinations of these. If an organization contemplates budget cuts, the effects of these cuts on employee attitudes can be studied with stratification by department, function, or region. Stratification ensures homogeneity within each stratum (that is, very few differences or dispersions on the variable of interest within each stratum), but heterogeneity (variability) between strata. In other words, there will be more between group differences than within-group differences.

- **Proportionate and Disproportionate Stratified Random Sampling:** Once the population has been stratified in some meaningful way, a sample of members from each stratum can be drawn using either a simple random sampling or a systematic sampling procedure. The subjects drawn from each stratum can be either proportionate or disproportionate to the number of elements in the stratum. For instance, if an organization employs 10 top managers, 30 middle managers, 50 lower-level managers, 100 supervisors, 500 clerks and 20 secretaries, and a stratified sample of about 140 people is needed for some specific survey, the researcher might decide to include in the sample 20% of members from each stratum. That is, members represented in the sample from each stratum will be proportionate to the total number of elements in the respective strata. This would mean that 2 from the top, 6 from the middle, and 10 from the lower levels of management will be included in the sample. In addition, 20 supervisors, 100 clerks, and 4 secretaries will be represented in the sample, as shown in the third column of Table given below. This type of sampling is called a proportionate stratified random sampling design.

  In situations like the one above, researchers might sometimes be concerned that information from only 2 members at the top and 6 from the middle levels would not truly reflect how all members at those levels would respond. Therefore, a researcher might decide instead, to use a disproportionate stratified random sampling procedure. The number of subjects from each stratum will now be altered, while keeping the sample size unchanged. Such a sampling design is illustrated in the far right-hand column in Table given below. The idea here is that the 60 clerks might be considered adequate to represent the population of 500 clerks; 7 out of 10 managers at the top level would also be considered representative of the top managers, and likewise 15 out of the 30 managers at the middle level. This redistribution of the numbers in the strata would be considered more appropriate and representative for the study than the previous proportionate sampling design.

  Disproportionate sampling decisions are made either when some stratum or strata are too small or too large, or when there is more variability suspected within a particular stratum. As an example, the educational levels among supervisors, which may be considered as influencing perceptions, may range from elementary school to master’s degrees. Here, more people will be sampled at the supervisory level. Disproportionate sampling is also sometimes done when it is easier, simpler, and less expensive to collect data from one or more strata than from others.

Prepared By:
Dr. Gaurav Sehgal
Associate Professor
School of Business Studies
Department of HRM & OB
Central University of Jammu, Jammu, J&K State

Proportionate and Disproportionate Stratified Random Sampling

Job Level	Number of Elements	Number of Subjects in the Sample: Proportionate Sampling (20% of the elements)	Disproportionate Sampling
Top management	10	2	7
Middle-level management	30	6	15
Lower-level management	50	10	20
Supervisors	100	20	30
Clerks	500	100	60
Secretaries	20	4	10
Total	710	142	142

In summary, stratified random sampling involves stratifying the elements along meaningful levels and taking proportionate or disproportionate samples from the strata. This sampling design is more efficient than the simple random sampling design because, for the same sample size, each important segment of the population is better represented, and more valuable and differentiated information is obtained with respect to each group.

(c) Cluster Sampling

Groups or chunks of elements that, ideally, would have heterogeneity among the members within each group are chosen for study in cluster sampling. This is in contrast to choosing some elements from the population as in simple random sampling, or stratifying and then choosing members from the strata as in stratified random sampling, or choosing every *n*th element in the population as in systematic sampling. When several groups with intragroup heterogeneity and intergroup homogeneity are found, then a random sampling of the clusters or groups can ideally be done and information gathered from each of the members in the randomly chosen clusters. Adhoc organizational committees drawn from various departments to offer inputs to the company president to enable him to make decisions on product development, budget allocations, marketing strategies, and the like, are good examples of different clusters. Each of these clusters or groups contains a heterogeneous collection of members with different interests, orientations, values, philosophy, and vested interests, drawn from different departments to offer a variety of perspectives. Drawing on their individual and combined insights, the president is able to make final decisions on strategic moves for the company. Cluster samples offer more heterogeneity within groups and more homogeneity among groups - the reverse of what we find in stratified random sampling, where there is homogeneity within each group and heterogeneity across groups.

The unit costs of cluster sampling are much lower than those of other probability sampling designs of simple or stratified random sampling or systematic sampling. However, cluster sampling exposes itself to greater biases and is the least generalizable of all the probability sampling designs, because most naturally occurring clusters in the organizational context do not contain heterogeneous elements. In other words, the conditions of intra-cluster heterogeneity and inter-cluster homogeneity are often not met. For these reasons, cluster sampling technique is not very common in

Prepared By:
Dr. Gaurav Sehgal
Associate Professor
School of Business Studies
Department of HRM & OB
Central University of Jammu, Jammu, J&K State

organizational research. Further, as in the case of the committee example cited above, duplication of members in several clusters is also possible. Moreover, for marketing research activities, naturally occurring clusters, such as, clusters of residents, buyers, students, or shops, do not have much heterogeneity among the elements. As stated earlier, there is more intra-cluster homogeneity than heterogeneity in such clusters. Hence, cluster sampling, though less costly, does not offer much efficiency in terms of precision or confidence in the results. However, cluster sampling offers convenience. For example, it is easier to inspect an assortment of units packed inside, say, four boxes (that is, all the elements in the four clusters) than to open 30 boxes in a shipment in order to inspect a few units from each at random.

- **Single-Stage and Multistage Cluster Sampling.** We have thus far discussed single-stage cluster sampling, which involves the division of the population into convenient clusters, randomly choosing the required number of clusters as sample subjects, and investigating all the elements in each of the randomly chosen clusters. Cluster sampling can also be done in several stages and is then known as multistage cluster sampling. For instance, if we were to do a national survey of the average monthly bank deposits, cluster sampling would first be used to select the urban, semi-urban, and rural geographical locations for study. At the next stage, particular areas in each of these locations would be chosen. At the third stage, banks within each area would be chosen. In other words, multistage cluster sampling involves a probability sampling of the primary sampling units; from each of these primary units, a probability sample of the secondary sampling units is then drawn; a third level of probability sampling is done from each of these secondary units, and so on, until we have reached the final stage of breakdown for the sample units, when we will sample every member in those units.

(d) Area Sampling

The area sampling design constitutes geographical clusters. That is, when the research pertains to populations within identifiable geographical areas, such as, counties, city blocks or particular boundaries within a locality, area sampling can be done. Thus, area sampling is a form of cluster sampling within an area. Sampling the needs of consumers before opening a 24-hour convenience store in a particular part of town would involve area sampling. Location plans for retail stores, advertisements focused specifically on local populations, and TV and radio programs beamed at specific areas could all use an area sampling design to gather information on the interests, attitudes, predispositions and behaviors of the local area people. Area sampling is less expensive than most other probability sampling designs, and it is not dependent on a population frame. A city map showing the blocks of the city would be adequate information to allow a researcher to take a sample of the blocks and obtain data from the residents therein.

(e) Double Sampling

This plan is resorted to when further information is needed from a subset of the group from which some information has already been collected for the same study. A sampling design where initially a sample is used in a study to collect some preliminary information of interest, and later a subsample of this primary sample is used to examine the matter in more detail, is called double sampling. For example, a structured interview might indicate that a subgroup of the respondents has more insight into the problems of the organization. These respondents might be interviewed again and asked additional questions. This research would have adopted a double sampling procedure.

3.6 Non-probability Sampling

In non-probability sampling designs, the elements in the population do not have any probabilities attached to their being chosen as sample subjects. This means that the findings from the study of the sample cannot be confidently generalized to the population. As stated earlier, however, researchers may at times be less concerned about generalizability than obtaining some preliminary information in a quick and inexpensive way. They would then resort to non-probability sampling. Sometimes nonprobability sampling could be the only way to obtain data.

Some of the non-probability sampling plans are more dependable than others and could offer some important leads to potentially useful information with regard to the population. The non-probability sampling designs, which fit into the broad categories of Convenience Sampling and Purposive Sampling.

In other words, there are two main types of non-probability sampling designs: Convenience Sampling and Purposive Sampling.

Convenience sampling is the least reliable of all sampling designs in terms of generalizability, but sometimes it may be the only viable alternative when quick and timely information is needed, or for exploratory research purposes.

Purposive sampling plans fall into two categories: Judgment and Quota Sampling designs.

Judgment sampling, though restricted in generalizability, may sometimes be the best sampling design choice, especially when there is a limited population that can supply the information needed. Quota sampling is often used on considerations of cost and time and the need to adequately represent minority elements in the population. Although the generalizability of all non-probability sampling designs is very restricted, they have certain advantages and are sometimes the only viable alternative for the researcher.

(a) Convenience Sampling: As its name implies, convenience sampling refers to the collection of information from members of the population who are conveniently available to provide it. One would expect that the - Pepsi Challenge contest was administered on a convenience sampling basis. Such a contest, with the purpose of determining whether people prefer one product to another, might be held at a shopping mall visited by many shoppers. Those inclined to take the test might form the sample for the study of how many people prefer Pepsi over Coke or product X to product Y. Such a sample is a convenience sample.

Consider another example: A convenience sample of five officers who attended the competitor‘s showcase demonstration at the county fair the previous evening offered the vice president of the company information on the - new products of the competitor and their pricing strategies, which helped the VP to formulate some ideas on the next steps to be taken by the company. Convenience sampling is most often used during the exploratory phase of a research project and is perhaps the best way of getting some basic information quickly and efficiently.

(b) Purposive Sampling: Instead of obtaining information from those who are most readily or conveniently available, it might sometimes become necessary to obtain information from specific target groups. The sampling here is confined to specific types of people who can provide the desired information, either because they are the only ones who have it, or conform to some criteria set by the researcher. This type of sampling design is called purposive sampling, and the two major types of purposive sampling - Judgment Sampling and Quota sampling.

- **Judgment Sampling**

  Judgment sampling involves the choice of subjects who are most advantageously placed or in the best position to provide the information required. For instance, if a researcher wants to find out what it takes for women managers to make it to the top, the only people who can give firsthand information are the women who have risen to the positions of presidents, vice presidents, and important top-level executives in work organizations. They could reasonably be expected to have expert knowledge by virtue of having gone through the experiences and processes themselves, and might perhaps be able to provide good data or information to the researcher. Thus, the judgment sampling design is used when a limited number or category of people have the information that is sought. In such cases, any type of probability sampling across a cross-section of the entire population is purposeless and not useful. Judgment sampling may curtail the generalizability of the findings, due to the fact that we are using a sample of experts who are conveniently available to us. However, it is the only viable sampling method for obtaining the type of information that is required from very specific pockets of people who alone possess the needed facts and can give the information sought. In organizational settings, and particularly for market research, opinion leaders who are very knowledgeable are included in the sample. Enlightened opinions, views, and knowledge constitute a rich data source. Judgment sampling calls for special efforts to locate and gain access to the individuals who do have the requisite information. As already stated this sampling design may be the only useful one for answering certain types of research questions.

- **Quota Sampling**

  Quota sampling, a second type of purposive sampling, ensures that certain groups are adequately represented in the study through the assignment of a quota. Generally, the quota fixed for each subgroup is based on the total numbers of each group in the population. However, since this is a non-probability sampling plan, the results are not generalizable to the population. Quota sampling can be considered as a form of proportionate stratified sampling, in which a predetermined proportion of people are sampled from different groups, but on a convenience basis. For instance, it may be surmised that the work attitude of blue-collar workers in an organization is quite different from that of white-collar workers. If there are 60% blue-collar workers and 40% white-collar workers in this organization, and if a total of 30 people are to be interviewed to find the answer to the research question, then a quota of 18 blue-collar workers and 12 white-collar workers will form the sample, because these numbers represent 60% and 40% of the sample size. The first 18 conveniently available blue-collar workers and 12 white-collar workers will be sampled according to this quota. Needless to say, the sample may not be totally representative of the population; hence the generalizability of the findings will be restricted. However, the convenience it offers in terms of effort, cost, and time makes quota sampling attractive for some research efforts. Quota sampling also becomes a necessity when a subset of the population is underrepresented in the organization - for example, minority groups, foremen, and so on. In other words, quota sampling ensures that all the subgroups in the population are adequately represented in the sample. Quota samples are basically stratified samples from which subjects are selected non-randomly.

In a workplace (and society) that is becoming increasingly heterogeneous because of the changing demographics, quota sampling can be expected to be used more frequently in the future. For example, quota sampling can be used to have some idea of the buying predispositions of various ethnic groups, for getting a feel of how employees from different nationalities perceive the organizational culture, and so on. Although quota sampling is not generalizable like stratified random sampling, it does offer some information, based on which further investigation, if necessary, can proceed. That is, it is possible that the first stage of research will use the nonprobability design of quota sampling, and once some useful information has been obtained, a probability design will follow. The converse is also entirely possible. A probability sampling design might indicate new areas for research, and non-probability sampling designs might be used to explore their feasibility.

3.7 Summary of Probability and Non-probability Sampling

Table below, summarizes the probability and non-probability sampling designs discussed so far, and their advantages and disadvantages.

Probability and Nonprobability Sampling Designs

Sampling Design	Description	Advantages	Disadvantages
Probability Sampling			
1. Simple random sampling	All elements in the population are considered and each element has an equal chance of being chosen as the subject.	High generalizability of findings.	Not as efficient as stratified sampling.
2. Systematic sampling	Every *n*th element in the population is chosen starting from a random point in the population frame.	Easy to use if population frame is available.	Systematic biases are possible.
3. Stratified random sampling (Str.R.S.) Proportionate Str.R.S.	Population is first divided into meaningful segments; thereafter subjects are drawn in proportion to their original numbers in the population.	Most efficient among all probability designs. All groups are adequately sampled and comparisons among groups are possible.	Stratification *must* be meaningful. More time-consuming than simple random sampling or systematic sampling.
Disproportionate Str.R.S.	Based on criteria other than their original population numbers.		Population frame for *each* stratum is essential.
4. Cluster sampling	Groups that have heterogeneous members are first identified; then some are chosen at random; all the members in each of the randomly chosen groups are studied.	In geographic clusters, costs of data collection are low.	The least reliable and efficient among all probablitiy sampling designs since subsets of clusters are more homogeneous than heterogeneous.
5. Area sampling	Cluster sampling within a particular area or locality.	Cost-effective. Useful for decisions relating to a particular location.	Takes time to collect data from an area.
6. Double sampling	The same sample or a subset of the sample is studied twice.	Offers more detailed information on the topic of study.	Original biases, if any, will be carried over. Individuals may not be happy responding a second time.
Nonprobability Sampling			
7. Convenience sampling	The most easily accessible members are chosen as subjects.	Quick, convenient, less expensive.	Not generalizable at all.
8. Judgment sampling	Subjects selected on the basis of their expertise in the subject investigated.	Sometimes, the only meaningful way to investigate.	Generalizability is questionable; not generalizable to entire population.
9. Quota sampling	Subjects are conveniently chosen from targeted groups according to some predetermined number or quota.	Very useful where minority participation in a study is critical.	Not easily generalizable.

3.8 Issues of Precision and Confidence in Determining Sample Size

Suppose we select 30 people from a population of 3,000 through a simple random sampling procedure. Will we be able to generalize our findings to the population with confidence, since we have chosen a probability design that has the most generalizability? What is the sample size that would be required to make reasonably precise generalizations with confidence? What do precision and confidence mean? These issues will be discussed in this section.

A reliable and valid sample should enable us to generalize the findings from the sample to the population under investigation. In other words, the sample statistics should be reliable estimates and reflect the population parameters as closely as possible within a narrow margin of error. No sample statistic ($\bar{X}$ for instance), is going to be *exactly* the same as the population parameter (μ), no matter how sophisticated the probability sampling design is. Remember that the very reason for a probability design is to increase the probability that the sample statistics will be as close as possible to the population parameters! Though the point estimate X may not accurately reflect the population mean μ, an interval estimate can be made within which μ will lie, with probabilities attached, that is, at particular confidence levels. The issues of confidence interval and confidence level are addressed in the following discussions on precision and confidence.

Prepared By:
Dr. Gaurav Sehgal
Associate Professor
School of Business Studies
Department of HRM & OB
Central University of Jammu, Jammu, J&K State

Precision: Precision refers to how close our estimate is to the true population characteristic. Usually, we would estimate the population parameter to fall within a range, based on the sample estimate. For example, let us say that from a study of a simple random sample of 50 of the total 300 employees in a workshop, we find that the average daily production rate per person is 50 pieces of a particular product ($X = 50$). We might then (by doing certain calculations, as we shall see later) be able to say that the *true* average daily production of the product (μ) would lie anywhere between 40 and 60 for the population of employees in the workshop.

In saying this, we offer an interval estimate, within which we expect the true population mean production to be ($\mu = 50 \pm 10$). The narrower this interval, the greater the precision. For instance, if we are able to estimate that the population mean would fall anywhere between 45 and 55 pieces of production ($\mu = 50 \pm 5$) rather than 40 and 60 ($\mu = 50 \pm 10$), then we would have more precision. That is, we would now estimate the mean to lie within a narrower range, which in turn means that we estimate with greater exactitude or precision. Precision is a function of the range of variability in the sampling distribution of the sample mean. That is, if we take a number of different samples from a population, and take the mean of each of these, we will usually find that they are all different, are normally distributed, and have a dispersion associated with them. The smaller this dispersion or variability, the greater the probability that the sample mean will be closer to the population mean. We need not necessarily take several different samples to estimate this variability. Even if we take only one sample of 30 subjects from the population, we will still be able to estimate the variability of the sampling distribution of the sample mean. This variability is called the standard error, denoted by $S_{\bar{x}}$. The standard error is calculated by the following formula:

$$S_X = \frac{S}{n}$$

where S_x is the standard deviation of the sample, n is the sample size, and $S_{\bar{x}}$, indicates the standard error or the extent of precision offered by the sample. Note that the standard error varies inversely with the square root of the sample size. Hence, if we want to reduce the standard error given a particular standard deviation in the sample, we need to increase the sample size. Another noteworthy point is that the smaller the variation in the population, the smaller the standard error, which in turn implies that the sample size need not be large. Thus, low variability in the population requires a smaller sample size. In sum, the closer we want our sample results to reflect the population characteristics, the greater will be the precision we would aim at. The greater the precision required, the larger is the sample size needed, especially when the variability in the population itself is large.

Confidence: Whereas precision denotes how close we estimate the population parameter based on the sample statistic, confidence denotes how *certain* we are that our estimates will really hold true for the population. In the previous example of production rate, we know we are more precise when we estimate the true mean production (μ) to fall somewhere between 45 and 55 pieces, than somewhere between 40 and 60. However, we may have more confidence in the latter estimation than in the former. After all, anyone can say with 100% certainty or confidence that the mean production (μ) will fall anywhere between zero and infinity! Other things being equal, the narrower the range, the lower the confidence. In other words, there is a trade-off between precision and confidence for any given sample size. In essence, confidence reflects the level of certainty with which we can state that our estimates of the population parameters, based on our sample statistics, will hold true. The level of confidence can range from 0 to 100%. A 95% confidence is the conventionally accepted level for most business research, most commonly expressed by denoting the significance level as $p = .05$. In other words, we say that at least 95 times out of 100, our estimate will reflect the true population characteristic.

3.9 Determining the Sample Size

Now that we are aware of the fact that the sample size is governed by the extent of precision and confidence desired, how do we determine the sample size required for our research? The procedure can be illustrated through an example.

Suppose a manager wants to be 95% confident that the expected monthly withdrawals in a bank will be within a confidence interval of ± \$500. Let us say that a study of a sample of clients indicates that the average withdrawals made by them have a standard deviation of \$3,500. What would be the sample size needed in this case?

We noted earlier that the population mean can be estimated by using the formula:

$$\mu = \bar{X} \pm K S_{\bar{x}}$$

Since the confidence level needed here is 95%, the applicable K value is 1.96 (t table). The interval estimate of ± \$500 will have to encompass a dispersion of (1.96 × standard error). That is,

$$500 = 1.96 \times S_{\bar{x}}$$
$$S_{\bar{x}} = 500/1.96 = 255.10$$

We already know that

$$S_{x} = \frac{s}{n}$$

$$255.10 = \frac{3500}{n}$$

$$n = 188$$

The sample size needed in the above was 188. Let us say that this bank has a total clientele of only 185. This means we cannot sample 188 clients. We can in this case apply the correction formula and see what sample size would be needed to have the same level of precision and confidence given the fact that we have a total of only 185 clients. The correction formula is as follows:

$$S_{\bar{x}} = \frac{S}{n} \times \frac{N-n}{N-1}$$

where N is the total number of elements in the population, n is the sample size to be estimated, $S_{\bar{x}}$ is the standard error of estimate of the mean, and S is the standard deviation of the sample mean.

Applying the correlation formula, we find that

$$255.10 = \frac{3500}{n} \times \frac{185-n}{184}$$

$$n = 94$$

We would now sample 94 of the total 185 clients.

To understand the impact of precision and/or confidence on the sample size, let us try changing the confidence level required in the bank withdrawal exercise which needed a sample size of 188 for a confidence level of 95%. Let us say that the bank manager now wants to be 99% sure that the expected monthly withdrawals will be within the interval of ±$500. What will be the sample size now needed?

$S_{\bar{x}}$ *will now be*

$$\frac{500}{2.576} = 194.099$$

$$194.099 = \frac{3500}{n}$$

$$n = 325$$

The sample has now to be increased 1.73 times (from 188 to 325) to increase the confidence level from 95% to 99%!

3.10 Importance of Sampling Design and Sample Size

It is now possible to see how both sampling design and the sample size are important to establish the representativeness of the sample for generalizability. If the appropriate sampling design is not used, a large sample size will not, in itself, allow the findings to be generalized to the population. Likewise, unless the sample size is adequate for the desired level of precision and confidence, no sampling design, however sophisticated, can be useful to the researcher in meeting the objectives of the study. Hence, sampling decisions should consider both the sampling design and the sample size. Too large a sample size, however (say, over 500) could also become a problem inasmuch as we would be prone to committing Type II errors. That is, we would accept the findings of our research, when in fact we should reject them. In other words, with too large a sample size, even weak relationships (say a correlation of .10 between two variables) might reach significance levels, and we would be inclined to believe that these significant relationships found in the sample are indeed true of the population, when in reality they may not be. Thus, neither too large nor too small sample sizes help research projects.

Another point to consider, even with the appropriate sample size, is whether statistical significance is more relevant than practical significance. For instance, a correlation of .25 may be statistically significant, but since this explains only about 6% of the variance (.252), how meaningful is it in terms of practical utility? Roscoe (1975) proposes the following rules of thumb for determining sample size:

1. Sample sizes larger than 30 and less than 500 are appropriate for most research.

Prepared By:
Dr. Gaurav Sehgal
Associate Professor
School of Business Studies
Department of HRM & OB
Central University of Jammu, Jammu, J&K State

2. Where samples are to be broken into subsamples; (male / females, juniors / seniors, etc.), a minimum sample size of 30 for each category is necessary.
3. In multivariate research (including multiple regression analyses), the sample size should be several times (preferably 10 times or more) as large as the number of variables in the study.
4. For simple experimental research with tight experimental controls (matched pairs, etc.), successful research is possible with samples as small as 10 to 20 in size.

3.11 Data Collection

Refer to: Chapter- 14
Book: "Fundamentals of Research Methodology and Statistics"
Author: "Yogesh Kumar Singh"
Publisher: "New Age International Publishers"