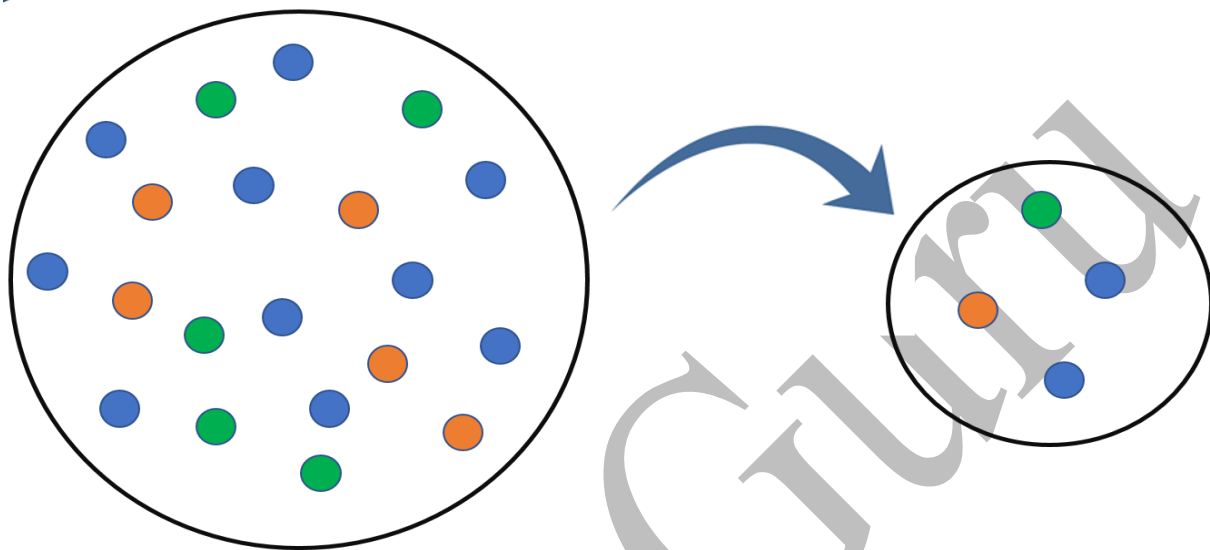# Sampling

Sampling Concept

Sampling is a method for collecting information and drawing inferences about a larger population or universe, from the analysis of only part thereof, the sample. In other words, a sample is a subset of the population that represents the entire group. When the population for universe) is too large for the researcher to survey all its members became of its cost, the member of personal to be employed, or the time constraint, a small carefully chosen sample is extracted to represent the whole. Sampling allows surveys of the complete population of a country, or subsections of it, to be carried out far more cheaply and frequently, and with resources devoted to improving the depth and quality of the information collected, in contrast with the shallow information obtainable from censuses.

Advantages of Sampling

- Sampling helps in substantial cost reduction of surveys

- Sample survey analyze the characteristics of a population

- Permits a high degree of accuracy due to a limited area of operations

- More detailed information can be collected through sample survey

- Collects the data required in far less time

- Intensive and exhaustive data are collected

- Study of samples involves less space and equipment

- Sampling is suitable when available sources are limited

Disadvantages of Sampling

- It may involve biased selection which leads to draw invalid conclusions

- Using a sample from a population involves risk

- Lack of representativeness of entire population

- Widely dispersion and heterogeneity of population may lead to difficulty in sampling

- Lack of specific and specialized knowledge in researcher

Essentials of a Good Sample

- A sample must represent a true picture of the population from which it is drawn.

- A sample must be unbiased by the sampling procedure.

- A sample must be taken at random so that every member of the population of data has an equal chance of selection.

- A sample must be sufficiently large but as economical as possible.

- A sample must be accurate and complete. It should not leave any information incomplete and should include all the respondents, units or teams included in the sample.

• Adequate sample size must be taken considering the degree of precision required in the results of inquiry.

In brief, a good sample should be truly representative in character. It should be selected at random and should be adequately proportional. These, in fact, are the attributes of a good sample.

Steps in Sampling Design

A sampling design is a definite plan for obtaining a sample from a given population. It refers to the technique or the procedure; the business researchers would adopt to select units for the sample. Sampling design is determined before data are collected.

1. Defining the Population

The first step is the most critical step wherein a researcher has to give an operational definition for his relevant population. Population in the most common form implies the total number of individuals in the area of study from whom information is being sought, but depending on the nature of study the population can be events, workplaces etc. The members who possess the characteristics or information required by the researcher constitute target population.

Defining the Population

Defining the Sample Unit

Determining the Sampling Frame

Selecting the Sampling Technique

Determining the Sample Size

Execution of Sampling Process

2. Defining the Sample Unit

A consequence to the above point is defining the sample unit. The sampling unit is the person, place or object about which or from which the information is required for research e.g. in case of study conducted on newspaper readership an entire household could be a sampling unit. However, in case of study on brand preference of college students in Jeanswear every college going student is a sampling unit.

3. Determining the Sampling Frame

Also referred to as 'Source List' it is a comprehensive listing of all the members of a population e.g. a telephone directory is a good sampling frame, as association directory e.g. NASSCOM will have a listing of all the firms in the industry. If the sampling frame is not available, then the researcher needs to prepare the list. It may be possible that the source list does not cover all the elements of population. In that case sampling frame error enters our research which has to be accounted for in further analysis.

4. Selecting the Sampling Technique

The researcher must decide on the type of sample i.e. a probability or non-probability based sampling techniques. The decision as regard to the technique to be used is affected by the objectives of study, the level of accuracy desired, time and cost etc. The different sampling techniques available to a researcher are discussed in detail in the next section.
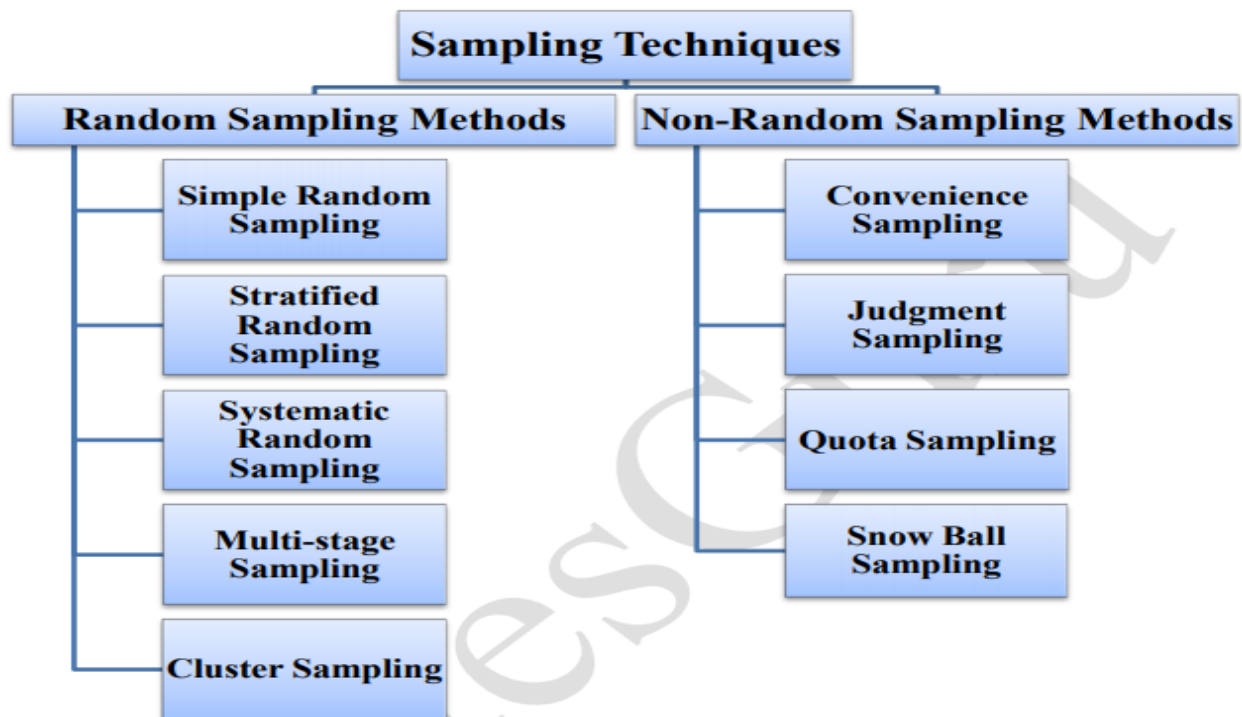
5. Determining the Sample Size

The million dollar question that faces the researcher is 'What sample size is needed?' A misconception is that a large sample size is a more representative sample. However there is no rigid rule that is used, although a formula does exist. A sample size is considered optimum if it fulfills the requirements of efficiency, representativeness, reliability and flexibility. Number of subjective factors like nature of population, nature of respondents, time and funds available, and sampling technique used etc. influence the decision on sample size.

6. Execution of Sampling Process

The last step involves working out the details of drawing a sample. Procedures for selecting each unit must be worked out. In case of non-response, the procedure to be adopted for filling the vacant unit should be stated clearly.

Types of Sampling Techniques

```
                    Sampling Techniques
        ┌───────────────────────┴──────────────────────────┐
Random Sampling Methods              Non-Random Sampling Methods
  │                                        │
  ├── Simple Random                        ├── Convenience
  │   Sampling                             │   Sampling
  │                                        │
  ├── Stratified                           ├── Judgment
  │   Random                               │   Sampling
  │   Sampling                             │
  │                                        ├── Quota Sampling
  ├── Systematic                           │
  │   Random                               ├── Snow Ball
  │   Sampling                             │   Sampling
  │
  ├── Multi-stage
  │   Sampling
  │
  └── Cluster Sampling
```

1. Random Sampling Methods

The first statistical sampling method is simple random sampling. In this method, each item in the population has the equal chances of being selected as part of the sample as any other item. For example, in a population of 1000 members, each of these members will have 1/1000 chances of being selected to be a part of a sample. It gets rid of bias in the population and gives a fair chance to all members to be included in the sample.

a. Simple Random Sampling: Every element has an equal chance of getting selected to be the part sample. It is used when we don't have any kind of prior information about the target population.

For example, in an organization of 500 employees, if the HR team decides on conducting team building activities, it is highly likely that they would prefer picking chits out of a bowl. In this case, each of the 500 employees has an equal opportunity of being selected.

b. Stratified Random Sampling: This technique divides the elements of the population into small subgroups (strata) based on the similarity in such a way that the elements within the group are homogeneous and heterogeneous among the other subgroups formed e.g. students of college can be divided into strata on the basis of gender, courses offered, age, etc. And then the elements are randomly selected from each of these strata. We need to have prior information about the population to create subgroups.

c. Systematic Random Sampling: Using systematic sampling method, members of a sample are chosen at regular intervals of a population. It requires selection of a starting point for the sample and sample size that can be repeated at regular intervals. This type of sampling method has a predefined interval and hence this sampling technique is the least time consuming.

For example, a researcher intends to collect a systematic sample of 500 people in a population of 5000. Each element of the population will be numbered from 1-5000 and every 10th individual will be chosen to be a part of the sample (Total population/ Sample Size = 5000/500 = 10).

d. Multi-stage Sampling: It is the combination of one or more methods described above. Population is divided into multiple clusters and then these clusters are further divided and grouped into various sub groups (strata) based on similarity. One or more clusters can be randomly selected from each stratum. This process continues until the cluster can't be divided anymore.

For example, country can be divided into states, cities, urban and rural and all the areas with similar characteristics can be merged together to form a strata.

e. Cluster Sampling: Cluster sampling is a method where the researchers divide the entire population into sections or clusters that represent a population. Clusters are identified and included in a sample on the basis of defining demographic parameters such as age, location, sex etc. which makes it extremely easy for a survey creator to derive effective inference from the feedback.

Cluster sampling can be done in following ways:

• Single Stage Cluster Sampling: When all elementary sampling units contained in one or more clusters, selected at random, are studied.

• Two Stage Cluster Sampling: Here first we randomly select clusters and then from those selected clusters we randomly select elements for sampling.

2. Non-Random Sampling Methods

The non-random sampling methods are often called non-probability sampling methods. In a nonrandom sampling method, the probability of any particular unit of the population being chosen is being unknown. It does not rely on randomization. This technique is more reliant on the researcher's ability to select elements for a sample. Outcome of sampling might be biased and makes difficult for all the elements of population to be part of the sample equally.

a. Convenience Sampling: This method is dependent on the ease of access to subjects such as surveying customers at a mall or passers-by on a busy street. In this the researcher has the freedom of choosing any respondent based on his convenience.

For example, in surveys conducted at a retail outlet or shopping mall interviews, people who happen to be at these places at the time the response was being taken become a part of the sample.

b. Judgment Sampling: Judgment Sampling is a purposive sampling where those respondents are deliberately made a part of sample, which meet research purpose based on researcher's own judgment.

For example, if a survey is being conducted on finding out what is takes to be a student union leader, then it is the people who are in this position i.e. student union leaders, or those who frequently interact with such people, who can give first hand information.

c. Quota Sampling: In Quota sampling, selection of members in this sampling technique happens on basis of a pre-set standard. In this case, as a sample is formed on basis of specific

attributes, the created sample will have the same attributes that are found in the total population. It is an extremely quick method of collecting samples.

d. Snow Ball Sampling: This technique is used in the situations where the population is completely unknown and rare. Therefore we will take the help from the first element which we select for the population and ask him to recommend other elements who will fit the description of the sample needed. So this referral technique goes on, increasing the size of population like a snowball.

For example, it will be extremely challenging to survey shelter less people or illegal immigrants. In such cases, using the snowball theory, researchers can track a few of that particular category to interview and results will be derived on that basis.

Determination of Sampling Size

A sample size is a part of the population chosen for a survey or experiment. For example, you might take a survey of dog owner's brand preferences. You won't want to survey all the millions of dog owners in the country (either because it's too expensive or time consuming), so you take a sample size. That may be several thousand owners. The sample size is a representation of all dog owners' brand preferences. If you choose your sample wisely, it will be a good representation.

In statistics, a sample refers to the observations drawn from a population. Sample size is used in research and defines the number of subjects that should be included within a sample. Having the right sample size is crucial in finding a statistically significant result. The larger the sample size, the more reliable the results; however, larger sample size means more time and money.

In order to do proper study of a problem, one must have proper sampling. In other words, sample size should be appropriate for analysis.

An important aspect of sampling is the decision on the size of sample i.e. the number of units to be selected from the population for carrying out the research. There is no key answer or solution to defining the most appropriate size. There are certain misconceptions regarding the size of sample like the sample should be 10% of the population or the sample size is proportional to the size of the universe. However as said earlier, these are only misconceptions.

The choice of the size of sample is affected by various factors discussed as below:

1. The Nature of Population

The degree of homogeneity or heterogeneity affects the size of a sample. If the population is homogeneous with respect to the characteristics of interest then even a small size of the sample is sufficient. However if the population is heterogeneous then a larger sample would be required to ensure adequate representativeness.

2. Nature of Respondent

If the respondents are easily available and accessible then required information can be obtained from a small sample. If, however, the respondents are uncooperative and non-response is expected to be high then a larger sample is required.

3. Nature of Study

A onetime study can be conducted using a large sample. In case of research studies which are of continuous nature and are to be intensively carried out, a small sample is more suitable as it is easy to manage and retain a small sample over a long span of time.

4. Sampling Technique Used

A very important factor influencing the size of sample is the sampling technique adopted. Firstly a non-probability technique requires a larger sample than a probability technique. Secondly within probability sampling, if simple random sampling is used it requires a larger sample than if stratification is used, where a small sample is sufficient.

5. Complexity of Tabulation

While deciding on the sample size the researcher should also consider the number of categories and classes into which the findings are to be grouped and analyzed. It has been seen that more the number of categories that are to be developed the larger is the sample size. Since each category should be adequately represented, a larger sample is required to give reliable measures of the smallest category.

6. Availability of Resources

The funds and the time available to a researcher influence the size of sample. Research is a time and money intensive task, with activities like preparation of instrument, hiring and training field staff, transportation costs etc. taking up a lot of resources. Hence if the researcher does not have enough time and funds available he will opt for a smaller sample.

7. Degree of Precision and Accuracy Required

It has become clear from our earlier discussion that precision, which is measured by standard error, will be high only if S.E is less or the sample size is large. Similarly to obtain a high degree of accuracy a larger sample is required.

Sources of Secondary Data

Secondary data are basically second-hand pieces of information. They are not gathered from the source as the primary data. To put it in other words, secondary data are those which are already collected. So they may be comparatively less reliable than the primary data. Secondary data is usually used when the time for the enquiry is compact and the exactness of the enquiry can be settled to an extent. However, secondary data can be gathered from different sources which can be categorized into 2 categories. Namely:

1. Published Sources:

Secondary data is usually gathered from the published (printed) sources. A few major sources of published information are mentioned below:

- Published articles of local bodies and Central and State Governments.

- Statistical synopses, census records and other reports issued by different departments of the Government.

- Official statements and publications of the foreign Governments.

- Publications and Reports of chambers of commerce, financial institutions, trade associations, etc.,

- Magazines, journals and periodicals.

- Publications of Government organizations like the Central Statistical Organization (CSO), National Sample Survey Organization (NSSO).

- Reports presented by Research Scholars, Bureaus, Economists, etc.,

2. Unpublished Sources:

Statistical data can be obtained from several unpublished references. Some of the major unpublished sources from which secondary data can be gathered are:

- The research works conducted by teachers, professors and professionals.

- The records that are maintained by private and business enterprises.

- Statistics maintained by different departments and agencies of the Central and State Governments, Undertakings, Corporations, etc.,