Chapter Seven: Sampling

Sampling refers to drawing a sample (subset) from a population (the fullest). The theory of sampling has been developed recently but this is not new. In our every day life we have been using sampling theory. In all cases we believe that the sample gives a correct idea about the population.

In research it also refers to choosing a smaller, more tangible number of people to take part in the research. The usual goal of sampling is to produce a representative sample (i.e. a sample that is similar to the population on all characteristics, except that it includes fewer people because it is a sample rather than a complete population).

The foremost purpose of sampling is to gather maximum information a bout the population under consideration at minimum cost, time and human power. This is best achieved when the sample contains all the properties of the population.

Terminology Used in Sampling

Here are some important terms used in sampling:

A sample: A set of element taken from a population (i.e. a finite subset of individuals defined in a population.)

Sampling unit**:** The constituents of a population, which are individual's to be sampled from the population and can not be further subdivided for the purpose of sampling at a time. Ex: to know the average income per family the head of the family is the sampling unit. To know the average yield of wheat each farmer owner's yield of wheat is the sampling unit.

Population: the full set of elements or people you are sampling. Technically known as Universe. The universe can be finite or infinite; in earlier case the number of items is certain.

Sampling Design: A definite plan for obtaining a sample from the sampling frame. It refers to technique / Procedure adopted by the researcher.

Census: A measurement of all the units in the population.

Statistics Vs Parameter

Statistics: a numerical characteristics of as ample (i.e. is a number that result from measuring all the units in the sample.)

Parameter: a numerical characteristic of a population (i.e., is a number that result from measuring all the units in the claimed population)

- Statistics derived from samples are used to estimate population

Sampling error Vs. Non- sampling error

Sampling error: Although the sample is the part of the population, it cannot be expected generally to supply full information about the population. So there may be in most cases difference between statistics and parameters. The discrepancy between a parameter and its estimate (statistics) due to sampling process is known as sampling error (i.e., the difference between the value of a sample statistics (such as the sample mean) and the true value of the population parameter (such as the population mean).

Non- sampling error: In all research / survey some errors may occur during collection of actual information. These errors are called non- sampling error.

Reasons for selecting sample: Why sample?

Sampling is inevitable in the following situations:

- Complete enumeration are practically impossible when the population is infinite,
- When the results are required in short time - it would be too time consuming to study the whole units in the population.
- When the area of survey is wide - it would to expensive and time consuming as will to study the whole units in the population
- When resource (Money, time and trained persons) are limited.
- When the items or units are destroyed under investigation for example a company to demonstrate that their car can survive certain crash tests. Obviously, the company cannot expected to crash every car, to see if it survives, the company crash only the sample of cars.

Principles of sampling

Samples have to provide good estimate. The following principle tells us that the sample methods provide such good estimate.

1. Principle of statistical regularity: A moderately large number of units chosen at random form large groups are almost sure on the average to possess the characteristics of the large group.
2. Principle of Inertia of large number: Other things being equal, as the sample size increase the result tend to be more accurate and reliable.
3. Principle of validity: This states that the sampling methods provide valid estimate a bout the population units (parameters).
4. Principle of Optimization: The principle takes in to account the desirability of obtaining a sampling design which gives optimum result. This minimizes the risk/ loss of the sampling design.

Types of sample Design

There are different types of sample design based on two factors viz. the representation basis and the Element Selection Technique. On the representation basis the sample may be probability sampling or it may be non- probability sampling. Probability sampling is based on the concept of random selection. Whereas non-probability is non-probability is non-rand sampling.

On the element selection basis the sample may be restricted or unrestricted. When each sample element is drawn individually from the population at large, then the sample so drawn is known as unrestricted sample, where as all other forms of sampling are covered under the term restricted sampling. However, Sample design is basically of two types: probability and non- probability sampling

Basics sample Design

Element selection Technique	Representation Basis	
	Probability	**Non-probability**
Unrestricted	Random Sampling (simple)	Haphazard/ convenience sampling
Restrict	Cluster sampling Systematic sampling Stratified sampling	Purposive sampling (Quota, judgment sampling)

Probability Sampling

A probability sampling is on where the selection of the units from the population is made according to known probability. The sample is based on probability theory. Every unit of the population of interest must have a know non-zero chance of being selected in to the sample.

- Assigns equal probability/ chance to each units of the population (Every element has equal (non zero) chance of being selected.
- Error of estimation/significance of results obtained can be measured.
- Best technique for representative sample (i.e., produce representative sample).
- Ensure the law of statistical regularity (i.e., on an average, sample chosen will have the same composition and structure as the universe/ population.)
- Even each combination (possible) of sample will have equal probability of being picked.
- All choices are independent of one another.

Types of probability sampling

(1) Simple Random sampling: A simple random sample from finite population is a sample selected such that each possible sample combination has equal probability of being chosen. It is also called unrestricted random sampling. In this type of probability sampling each unit in the population is identified, and each unit has an equal chance of being in the sample. The selection of each unit is independent of the

selection of every other unit. Selection of one unit doesn't affect the chance of any other unit

Methods of selection of simple random sampling

(a) Lottery method: This is the most simple and popular method. In this method all the items of the population are assigned a number on a separate slip of paper of same size, shape and color. They are folded and mixed in a container. The required number of slip is selected at random for the desire sample size. If the universe/population is infinite, this method is inapplicable.

(b) Table of Random Numbers: As the lottery method cannot be used, when the population is infinite, the alternative method is that of using the table of random numbers. A random number table is so constructed that all digits 0 to 9 appear independent of each other with equal frequency. If we have to select sample from population of size N= 100, then the numbers from 001 to 100.

Procedures: Units of the population from which a sample is required are assigned with equal number of digits. When the size of the population is less than thousand, three digit numbers 000, 001, 002, --- 999 are assigned. We may start at any place and may go on in any direction such as a column wise or row-wise in a random number table. But consecutive numbers are to be used. If any random number is greater than the population size N the N can be subtracted from the random number drawn. This can be repeatedly until the number is less than N or equal to N.

Example: In an area there are 500 families. Using the following extract from a table of random number select a simple of 15 families to find out the standard of living of the families in that area.

4652	3819	8431	2150	2352	2472	0043	3488
9031	7617	1220	4129	7148	1943	4890	1749
2030	2327	7353	6007	9410	9179	2722	8445
0641	1489	0328	0385	8488	0422	7209	4950

In the above random number table we can start from any row/column and read three digit numbers continuously row-wise or column-wise.

Now we start from the third row, the numbers are:

203	023	277	353	600	794	109
179	272	284	450	641	148	908
280.						

Since some numbers are greater than 500 we subtract 500 from those numbers and we rewrite the selected numbers as follow.

203	023	277	353	100	294	109
179	272	284	450	641	141	408
280.						

(2) Systematic Random sampling: each unit in the population is identified and each unit has an equal chance of being in the sample. It is also called Quasi- random sampling. Selection procedures: systematic sampling involves three steps:

1 st: Determine the sampling interval, which is symbolized by "K" (i.e., it is the population size divided by the desired sample size).

2 nd: Randomly select a number between 1 and k and include that person/unit in your sample.

3 rd: Include all kth elements in your sample. For example if K is10 and your random selected number between 1 and 10, for instance 5, then you will select persons 5,15,25----etc. When you get to the end of your sampling frame you will have all the people to be included in your sample.

Stratified Random sampling :

All of the methods of sampling the procedure commonly used in surveys are stratified sampling. This technique is mainly used to reduce the population Heterogeneity and to increase the efficiency of estimate.

Stratification means division in to groups. In this method the population is divided in to a number of sub-groups/strata. The strata should be so formed that each stratum is homogeneous as far as possible. Then from each stratum a simple random sample may be selected and these are combined together to form the

required sample from the population. There are two types of stratified sampling. They are Proportional and Non- proportional. In the proportional sampling equal and proportional representation is given to subgroups or strata. If the number of items is large the sample will have a higher size and vice versa. The population size denoted by N and the sample size is denoted by 'n' the sample size is allocated to each stratum. That is given by n/N =C. so in this method each stratum is according to its size.

In Non-proportional sample, equal representation is given to all the sub-strata regardless of their existence in the population.

Example: A sample of 50 students is to be drawn from a population consisting of 500 students belonging to two institutions A and B. The number of student in the institution A is 200 and the institution B is 300. How will you draw the sample using proportional allocation? There are two strata in this case with sizes N_1=200 and N_2 =300 and the total population N= N_1 + N_2 = 500.

The sample size is 50.

If n_1 and n_2 are the sample size:

$$n_1 = \frac{n \times N1}{N} = \frac{50 \times 200}{500} = 20$$

$$n_2 = \frac{n \times N_2}{N} = \frac{50 \times 300}{500} = 30$$

The sample size is 20 from A and 30 from B. then the unit from each institution is to be selected by simple random sampling.

---

(3) Cluster Sampling: cluster sampling views the units in the population as not only being members of the total population but as members also of naturally occurring in

a cluster within the population. Cluster sampling is used in a large geographic sample where No list is available of all units in the population but the population boundaries can be well defined.

Example: To obtain information about the drug habit of all high school students in a state:

You could obtain the list of all school districts in the state and select a set of school districts.

↓

Then with in each selected school district list all the high schools and select a SRS of classes.

↓

Within each selected high school list all the high school of classes

↓

Then use the high school students in that class as your sample.

Cluster sampling must use a random sampling method at each stage. Generally it divides area of interest in to a number of smaller non-overlapping areas /cluster.

Non- Probability/ sampling

It is the one where discretion is used to select representative units from the population (or) to infer that a sample is representative of the population. This method is also called Non-Random/judgment / purposive sampling. It mainly used for opinion survey.

Focus on volunteers, easily available units, or those just happen to be present when the researcher is done. It is useful for quick and cheap studies, case studies, for qualitative research, for pilot studies, and for developing hypotheses for further research.

There are FOUR main types of Non-probability methods/procedures:

(1) Convenience sampling: also called accidental /Haphazard /man in the street sampling. The researcher selects units that are convenient, close at hand, and easy to reach.

(2) Purposive sampling: The researchers select the units with some purpose in mind. The researcher specifies the characteristics of the population of interest and then locates individuals who match those characteristics. Ex: Experts on coops, students who lives in dorm on campus.

(3) Quota sampling: The researcher constructs quotas for different types of units. It also often involves setting quotas and then use convenience sampling to obtain those Quotas.

(4) Snowball sampling: Each research participant is asked to identify other potential research participants who have a certain characteristics. You start with one or few participants, ask them for more, find those, ask them for some and continue until you have a sufficient sample size. It usually used where no sampling frame exist.
