SAMPLING

Dr. A. B. Patankar (Athawale)

When secondary data are not available for the problem under study, a decision may be taken to collect primary data. These data may be obtained by following either population method (census method) or sample method.

POPULATION METHOD OR CENSUS METHOD

Population represents the whole area. Under the census or complete enumeration survey method, data are collected for each and every unit of the population or universe which is the complete set of items.

For example, if the average wages of workers in sugar industry in India is to be calculated, then it would be obtained from each and every worker working in the sugar industry. Then it is dividing the total wages received by the number of workers working in sugar industry. It is the average wages.

According to Simpson and Kafka, "Population or a universe means an aggregate of items possessing a common trait or traits".

The population can be classified into two types: They are (a) Finite or Infinite (b) Real or Hypothetical.

Finite or Infinite Population

If the number of elements can be counted in the population, then it is called finite population.

For example, number of students in a college, number of people in a village, number of schools in a city. If the number of elements cannot be counted in the population, then it is called infinite population. For example, number of stars in the sky, number of viewers in TV Programmes, number of readers in a newspaper.

Real or Hypothetical Population

In real population the elements or items really exist.

For example, number of factories in a district, number of people in a city. In hypothetical population, elements may not really exist; it is also known as theoretical population. For example, tossing a coin or throwing a dice.

Merits of Census Method

Following are the merits of census method:

1. Data are obtained from each and every unit of the population.
2. Data results are more representative, accurate and reliable.
3. Data of complete enumeration census can be widely used as a basis for various surveys.
4. The characters of all the limits of the population can be studied.
5. Intensive study is possible.

Demerits

The following are the demerits of census method:

1. It requires more time and cost.
2. It is not required for all types of research.
3. Inefficient and inexperience researcher may collect wrong information.

SAMPLING

Sampling is the process of learning about the population on the basis of a sample drawn from it. Thus, in the sampling technique instead of every unit of the universe, only a part of the universe is studied and the conclusions are drawn on that basis for the entire universe.

A sample is a subset of population units. The process of sampling involves three elements:

1. Selecting the sample
2. Collecting the information and
3. Making an inference about the population

The basic objective of sample study is to draw inference about the population. In other words, sampling is a tool, which helps to know the characteristics of the universe or population by examining only a small part of it.

Merits of Sample Method

The following are the merits of sample method.

1. The organisation and administration of sample surveys are easy.
2. It reduces cost, time and energy.
3. It gives accurate result.
4. It provides for detailed enquiry.

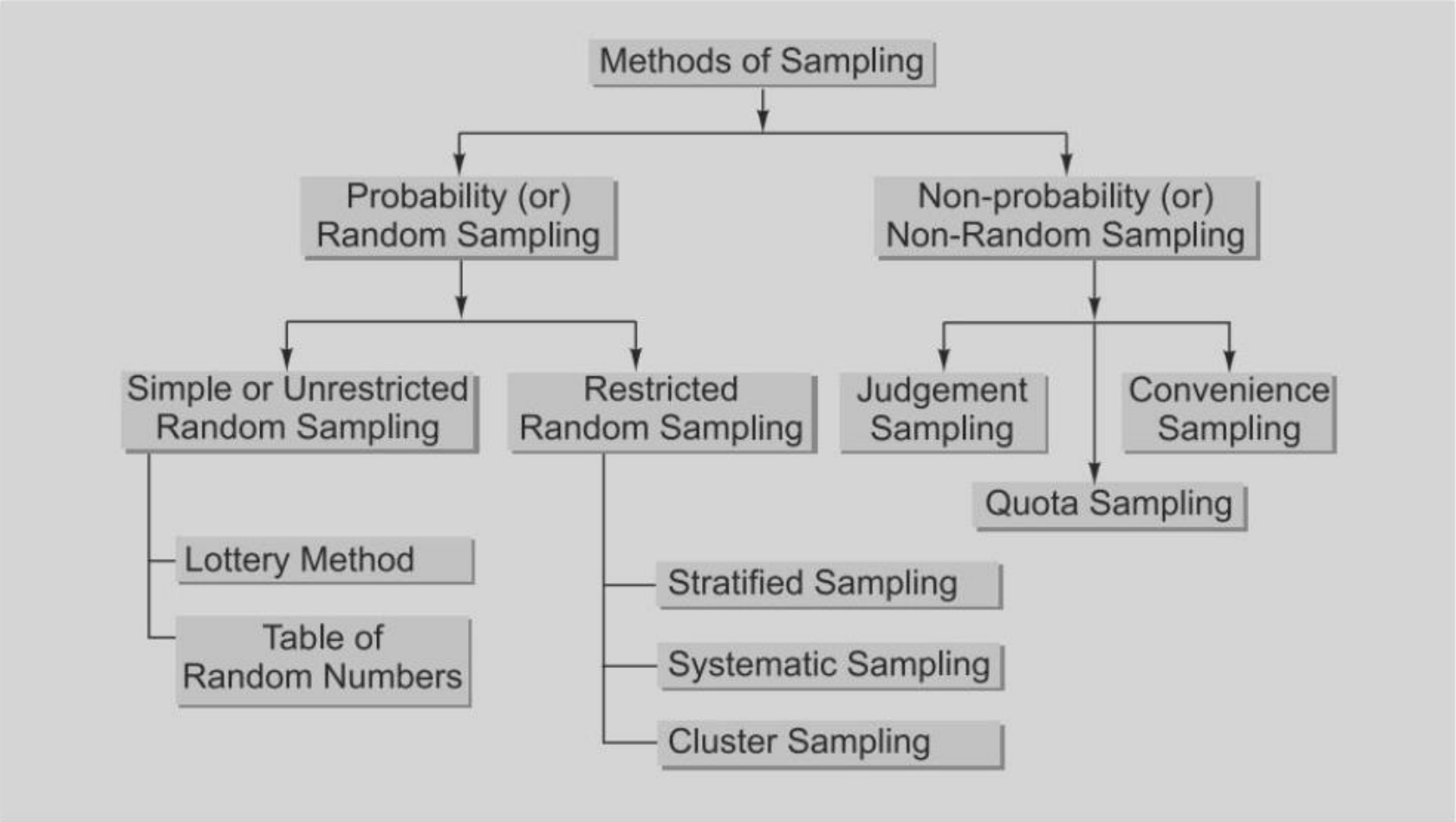
Demerits of Sample Method

Following are the demerits of sample method:

1. Data are not obtained from each and every unit of the population.
2. Reliable results are not possible.

METHODS OF SAMPLING

Statistical data are collected under sampling through the various methods. They are tabulated as follows:



Probability (or) Random Sampling

It is a method in which each item of the universe has the equal chance of being selected. This implies that the selection of sample items is independence of the person making the study.

Random Sampling can be classified into simple or unrestricted random sampling and restricted random sampling.

(a) Simple (or) Unrestricted Random Sampling It is a technique in which sample is so drawn that each and every unit of the population has an equal and independent chance of being included in the sample.

Under simple random sampling, data are collected through lottery method and table of random numbers.

(i) Lottery Method Under this method, all the items of the universe are numbered on separate slips of paper of same size and colour. These slips are folded and put in a box or containers. Then the required number of slips are selected from the box or container.

For example, if we want to take a sample of 10 persons out of a population of 100, the procedure is to write the names of the 100 persons on separate slips of paper, fold those slips, mix them thoroughly and then make a blindfold selection of 10 slips.

(ii) Table of Random Numbers Random number table is commonly used to select the sample, when the size of the population is very large. It is a table of digits which have been generated by a random process.

Several standard tables of random numbers are available, among which the following may be specially mentioned as they have been tested extensively for randomness.

1. Tippett's (1927) random number tables consisting of 41600 random digits grouped into 10,400 sets of four-digit random numbers.
2. Fisher and Yates (1938) table of random numbers consisting of 15000 random digits arranged into 1500 sets of 10-digit random numbers.
3. Kendal and Babington Smith (1939) table of random numbers consisting of 1,00,000 digits grouped into 25000 sets of four-digit random numbers.
4. Rand Corporation (1955) table of random numbers consisting of 1,00,000 random digitsgrouped into 20,000 sets of five-digit random numbers.

It is important that the starting point in the table of random numbers be selected in some random fashion so that every unit has an equal chance of being selected.

Merits of Simple Random Sampling It has the following advantages:

1. There is no possibility of personal bias since the selection of items in a sample depends entirely on chance.
2. It represents the universe in a better way. As the size of the sample increases, it becomes increasingly representative of the population.
3. It provides the most reliable information at the least cost.

Limitations of Simple Random Sampling This method is however associated with the following limitations:

1. It may produce the most non-random looking results.
2. The size of the sample is usually larger under random sampling than stratified sampling.
3. Time and cost of collecting data become too large.

Restricted Random Sampling

(i) Stratified Sampling Stratified random sampling (or) simply stratified sampling is one of the random methods used to design a more efficient sample than obtained by the simple random procedure. In stratified random sampling, the sampling is designed so that a designated number of items is chosen from each stratum.

The following are the procedure to be followed while applying stratified random sampling technique:

1. The universe to be sampled is subdivided into groups which are mutually exclusive and include all items in the universe.
2. A simple random sample is then chosen independently from each group.

Proportional and Disproportional Stratified Sample

In a proportional stratified sampling plan, the number of items drawn from each strata is proportional to the size of the strata. For example, if the population is divided into five groups, their respective sizes being 10, 15, 20, 30 and 25 per cent of the population and a sample of 5,000 is drawn, the desired proportional sample may be obtained in the following manner:

From stratum one	5,000 (0.10)	=	500	items
From stratum two	5,000 (0.15)	=	750	items
From stratum three	5,000 (0.20)	=	1,000	items
From stratum four	5,000 (0.30)	=	1,500	items
From stratum five	5,000 (0.25)	=	1,250	items
	Total	=	5,000	items

Proportional stratification yields a sample that represents the universe with respect to the proportion in each stratum in the population. This procedure is satisfactory if there is no great difference in dispersion from stratum to stratum. But it is certainly not the most efficient procedure, especially when there is considerable variation in different strata. This indicates that in order to obtain maximum efficiency in stratification,, we should assign greater representation to a stratum with a large dispersion and smaller representation to one with small variation.

In *disproportional stratified sampling* an equal number of cases is taken from each stratum regardless of how the stratum is represented in the universe. Thus, in the above example, an equal number of items (1,000) from each stratum may be drawn. In practice disproportional sampling is common when sampling forms a highly variable universe,, wherein the variation of the measurements differs greatly from stratum to stratum.

Illustration 1. You are given the following data of the number of lecturers, readers and professors in a University:

Length of service	*Lecturers*	*Readers*	*Professors*	*Total*
Less than 5 yrs	2,000	250	50	2,300
5–10 yrs	3,000	220	80	3,300
10–15 yrs	1,500	170	30	1,700
More than 15 yrs	880	80	40	1,000
Total	7,380	720	200	8302

Work out how many lecturers, readers and professors would be selected from each category if *(i)* we follow stratified proportionate sampling method and take 10% of the universe equivalent to the sample size, *(ii)* if the size of the sample is 10% of the universe but the lecturers, readers and professors are to be in the ratio of 5 : 3 : 2 and weightage of the length of service is to be in the ratio of 4 : 3 : 2 : 1.

Solution. *(i)* The sample size is 10% of the universe hence 830 persons would be selected in the sample. Since 12 strata are formed and we want to follow proportionate stratified sampling method, we will take 10% from each stratum. The number of persons selected shall be as follows:

Length of service	*Lecturers*	*Readers*	*Professors*	*Total*
Less than 5 yrs.	200	25	5	230
5-10 yrs.	300	22	8	330
10-15 yrs.	150	17	3	170
More than 15 yrs.	88	8	4	100
Total	738	72	20	830

(ii) In the second case also the size of sample is 830 but the lecturers, readers and professors are to be in the ratio of 5 : 3 : 2 of the sample, *i.e.*, we take $\frac{830 \times 5}{10} = 415$ lecturers; $\frac{830 \times 3}{10} = 249$ readers, and $\frac{830 \times 2}{10} = 166$ professors. Since the weightage to length of service is 4 : 3 : 2 : 1, the number selected from each category shall be as given in the table below:

Length of service	*Lecturers*	*Readers*	*Professors*	*Total*
Less than 5 years	$\frac{415 \times 4}{10} = 166$ or 124	$\frac{249 \times 4}{10} = 99.6$ or 100	$\frac{166 \times 4}{10} = 66.4$ or 66	332
5—10 years	$\frac{415 \times 3}{10} = 124.5$ or 124	$\frac{249 \times 3}{10} = 74.7$ or 74	$\frac{166 \times 3}{10} = 49.80$ or 50	248
10—15 years	$\frac{415 \times 2}{10} = 83$	$\frac{249 \times 2}{10} = 49.8$ or 50	$\frac{166 \times 2}{10} = 33.2$ or 33	166
Above 15 years	$\frac{415 \times 1}{10} = 41.5$ or 42	$\frac{249 \times 1}{10} = 24.9$ or 25	$\frac{166 \times 1}{10} = 16.6$ or 17	84
Total	415	249	166	830

Merits Stratified sampling methods have the following advantages:

1. A more representative sample is secured.
2. Stratified sampling ensures greater accuracy.
3. Stratified samples can be more concentrated geographically.

Demerits Demerits of this method are:

1. The sample may have the effect of bias if proper stratification of the population is not done.
2. In the absence of skilled sampling supervisors, a random selection within each stratum may not be ensured.
3. Cost of observation may be quite high.

(ii) Systematic Sampling (or) Quasi-Random Sampling This method is popularly used in those cases where a complete cost of the population from which sample to be drawn is available. The list may be prepared in alphabetical, geographical, numerical or some other order. The items are serially numbered.

It is relatively a simple technique and may be more efficient statistically. In this method, the first item is selected at random generally by following the lottery method. Subsequent items are selected by taking every K^{th} item from the list, where K refers to the sampling interval or sampling ratio symbolically.

$$K = \frac{N}{n}$$

where, K = Sampling Interval

N = Universe Size

n = Sample Size

While calculating k, it is possible that we get a fractional value. In such a case we should use approximation procedure, *i.e.*, if the fraction is less than 0.5 it should be omitted and if it is more than 0.5 it should be taken as 1. If it is exactly 0.5 it should be omitted, if the number is even and should be taken as 1, if the number is odd. This is based on the principle that the number after approximation should preferably be even. For example, if the number of students is respectively 1,020, 1,150 and 1,100 and we want to take a sample of 200, k shall be:

(i) $$k = \frac{1020}{200} = 5.1 \text{ or } 5$$

(ii) $$k = \frac{1150}{200} = 5.75 \text{ or } 6$$

(iii) $$k = \frac{1100}{200} = 5.5 \text{ or } 6.$$

Illustration 2. In a class there are 96 students with Roll Nos. from 1 to 96. It is desired to take sample of 10 students. Use the systematic sampling method to determine the sample size.

Solution.

$$k = \frac{N}{n} = \frac{96}{10} = 9.6 \text{ or } 10.$$

From 1 to 96 Roll Nos. the first students between 1 and k, *i.e.*, 1 and 10, will be selected at random and then we will go on taking every *kth* student. Suppose the first student comes out to be 4th. The sample would then consist of the following Roll Nos.

4, 14, 24, 34, 44, 54, 64, 74, 84, 94

Systematic sampling is relatively a simple technique and may be more efficient statistically than simple random sampling provided the lists are arranged wholly at random. However, it is rarely that this requirement is fulfilled. The nearest approach to randomness is provided by alphabetical lists such as are found in telephone directory although even these may have certain non-random characteristics.

Merits The merits of systematic sampling are:

- The systematic sampling design is simple and convenient to adopt.
- The time and work involved in sampling are relatively less.

Demerits The limitations of systematic sampling are:

1. It is less representative if we are dealing with populations having 'hidden periodicities'.
2. Personal bias of investigators.

(iii) Multi-Stage Sampling (or) Cluster Sampling Under this method, the random selection is made of primary, intermediate and final units from a given population or stratum. It is a method of sampling which is carried out in several stages.

For example, if we want to take a sample of 5000 households from Tamil Nadu state, first, the state may be divided into number of districts and a few districts may be subdivided into a number of villages and a sample of villages may be taken randomly. At the third stage, a number of households may be selected from each of the villages selected at the second stage.

Merits Cluster sampling has its own advantages. They are:

1. It introduces flexibility in the sampling method.
2. Large area may be covered.

Demerits The demerits of cluster sampling are:

1. It is less accurate than a sample containing the same number of final stage units which have been selected by some single stage process.
2. It is a long process.

Non-Probability Sampling Methods

Non-Probability sampling methods can be classified into three types. They are:

(a) Judgement Sampling

(b) Quota Sampling

(c) Convenience Sampling

(a) Judgement Sampling (or) Purposive (or) Deliberate Sampling In this method of sampling, the choice of sample items depends exclusively on the judgement of the investigator. In other words, the investigator exercises his judgement in the choice and includes those items in the sample, which is most typical of the universe with respect to its characteristics under investigation.

For example, if a sample of 10 students is to be selected from 60, for analysing the habits of students, the investigator would select 10 students, who in his opinion, are representative of the class.

(b) Quota Sampling In non-probability category, it is a commonly used sampling technique. It is like the stratified sampling. The population is divided into various quotas with respect to some common character. Then the required sample is to be selected from these quota.

Quota sampling is often used in public opinion studies. In quota sampling, the sampling within each cell is not done at random, the field representatives are given wide latitude in the selection of respondents to meet their quotas.

If the interviewers are carefully trained and if they follow their instructions closely, then satisfactory results can be acheived.

(c) Convenience Sampling A convenience sample is obtained by selecting 'convenient' population limits. The method of convenience sampling is also called the chunk. A chunk refers to that fraction of the population being investigated which is selected neither by probability nor by judgement but by convenience.

A sample may be obtained from readily available lists such as automobile registrations, telephone directories etc., is a convenience sample and not a random sample even if the sample is drawn at random from the lists. However, convenience sampling is often used for making pilot studies. Questions may be tested and preliminary information may be obtained by the chunk before the final sampling design is decided upon.