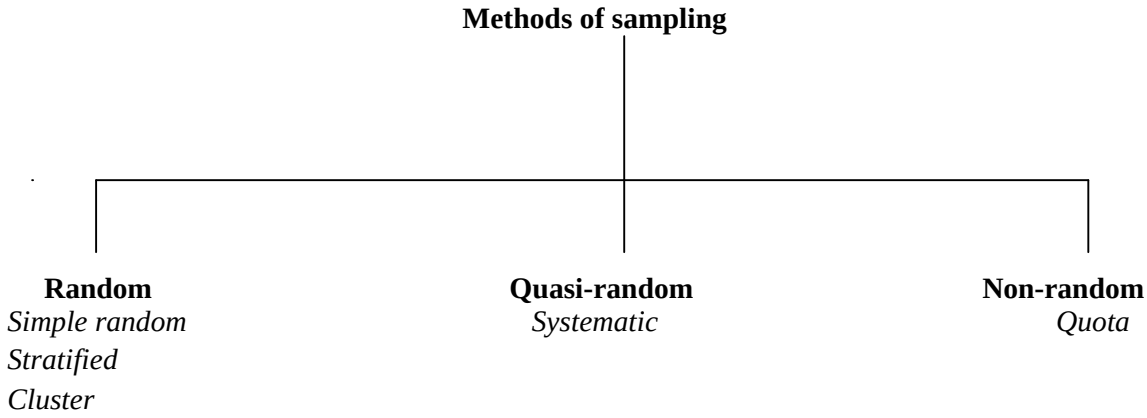


**Fig. 1**

**Fig. 1** shows the some of the different methods of sampling belonging to the three categories.

For random sampling, a sampling frame is not only a prerequisite but it also has to satisfy the following qualities:

- Completeness (all items of the population must be listed)
- Accuracy (The collected information must be correct and relevant to the theme of the survey)
- Adequacy (The sample should be fully representative of the population)
- Up-to-dateness (There should be no missing items)
- Convenience (The data should be readily accessible)
- Non-duplication (No bit of information should appear twice)

**Note**: Random data can also be obtained by the use of the random number tables.

**Simple random sampling**

The feature of this method of sampling is that every item in the population must be equally likely to be selected in the sample.

*Merit*

- In most cases, it is unbiased and hence the best method of sampling

*Limitations*

- The selected items are subject to the full range of variation inherent in the population
- An unrepresentative sample may result
- The items may be scattered throughout the population, making it expensive and time-consuming to collect the data
- An adequate sampling frame may not exist
- If information is lost, there is no guarantee of retrieving identical data by using random sampling (as opposed to cluster or quota sampling)

**Systematic (periodic) sampling**

In this method of sampling, the items are first arranged in some order (ascending or descending order with respect to some factor like age, height, weight). It may be easier to assign a number to each individual. The first individual is chosen at random and subsequent individuals are then selected at a regular interval known as the *period*. The period is determined according to the population and sample sizes. If the population size is $N$ and the sample size is $n$, then the period is $\frac{N}{n}$ (or the nearest integer to the value of $\frac{N}{n}$).

**Example**

It is required to select a sample of 12 individuals from a population of 96 by using systematic sampling. (Assume that the first individual selected is the 17th.)

*Solution*

Since $N = 96$ and $n = 12$, the period is 8. The individuals selected will have ranks

17 25 33 41 49 57 65 73 81 89 (97)

As 97 lies outside the range of the population numbers, we start subtracting 8 from the rank of the first selected observation to obtain the remaining individuals

9 1 (–7)

–7 also lies outside the range (therefore we stop selection).

*Merits*

- It is quick and easy given a sampling frame
- It is reasonably random provided there is no pattern in the population

*Limitations*

- It requires a sampling frame
- It requires direct access to every individual in the population
- Only combinations of items can be chosen (if the 2nd individual is chosen, then the 3rd has no chance of being in the sample)
- If the period of selection is a multiple of some *hidden periodicity* in the population, the sample will be heavily biased.

The concept of *hidden periodicity*

**Example**

The municipality of a certain town decides to carry out a survey in order to check whether the street lighting system is working well. The investigator opts for systematic sampling with a period of 20 houses. He therefore selects flats at an interval of 20 from blocks of flats in a certain area. It so happens that all the blocks contain 10 flats and that the investigator chooses the first flat number to be the $11^{th}$.

It is clear that the houses found in the corner of the blocks will have better
street lighting since they will receive light from the main street and the minor one. The period in the population of houses is 10 and the period of selection is 20 – thus, every selected house ($11^{th}$, $31^{st}$, $51^{st}$, … ) will be found in the corner of its block because the first one selected is the $11^{th}$. The investigator phones the owners of the chosen houses and asks them about their opinion on the street lighting system. It is obvious that the probability of them not receiving adequate lighting will be very small. Thus, the sample is heavily biased.

**Stratified (random) sampling**

This type of sampling is used when the population is divided into layers, classes or *strata*. The feature of stratified sampling is that any layer should be proportionately represented in the sample, that is, the percentage represented by a layer in the population and the sample should be identical. After determining the number of items to be chosen from each *stratum*, these items are chosen *at random* from their respective stratum.

**Example**

A survey is carried out in order to investigate about the opinion of pupils and students about the education system in a country. Assuming that the population consists of 30000 primary, 15000 secondary and 5000 tertiary students, what is the composition of a stratified random sample of 120 students?

*Solution*

The percentages for primary, secondary and tertiary students in the population are easily calculated as 60%, 30% and 10% respectively. Thus, the sample should consist of 72 primary (60% of 120), 36 (30% of 120) secondary and 12 (10% of 120) tertiary students respectively.

*Merits*

- It ensures proportionate representation of the different classes of the population
- The sample is fully representative of the population
- Since the items are selected at random, inferences can be made about each stratum
- The accuracy of *sample statistics* is increased.

*Limitations*

- It is very expensive to collect information about each stratum

**Quota sampling**

This is very widely used in market research. The population is divided into groups in terms of age, gender, income level, etc, … (depending of what is required in the survey). The interviewer is told how many people to interview but he has no instructions whatsoever about how to locate them (he can do so at his own convenience). He simply has to fulfil the quota (number of respondents).

*Advantages*

1. It is cheap and administratively easy.
2. A much larger sample can be selected.
3. No sampling frame is necessary.
4. It may sometimes be the only possible approach (e. g television audience marketing)
5. With proper training and briefing, accurate data can be obtained.

*Drawbacks*

1. It is biased.
2. It rules out any valid estimate of the sampling error.

**Cluster sampling**

Natural subgroupings of a population are called clusters. If we consider, e. g, children going to state primary schools from which to select a sample, then these schools themselves are natural clusters. In cluster sampling, we choose a random sample of clusters and then select members randomly within the selected clusters. Each cluster should be as similar as possible to another.

*Advantages*

1. No complete sampling frame is necessary.
2. Far less expensive since little organisation or structure is involved.

*Drawbacks*

1. It has the potential of considerable bias.
2. It does not give a precise picture of the whole population.