***Population:** The population is the entire group that you want to draw conclusions about.

***Sample:** The sample is the specific group of individuals that you will collect data from.

***Sampling frame**: The sampling frame is the actual list of individuals that the sample will be drawn from. Ideally, it should include the entire target population (and nobody who is not part of that population).

For example: You are doing research on working conditions at Company X. Your population is all 1000 employees of the company. Your sampling frame is the company's HR database which lists the names and contact details of every employee.

***Sample size:** The sample size is defined as the number of observations used for determining the estimations of a given population. The size of the sample has been drawn from the population. Sampling is the process of selection of a subset of individuals from the population to estimate the characteristics of the whole population.

Sampling method

When we conduct research about a group of people, it's rarely possible to collect data from every person in that group. Instead, we select a sample. The sample is the group of individuals who will actually participate in the research.

To draw valid conclusions from our results, we have to carefully decide how we will select a sample that is representative of the group as a whole. There are two types of sampling methods:

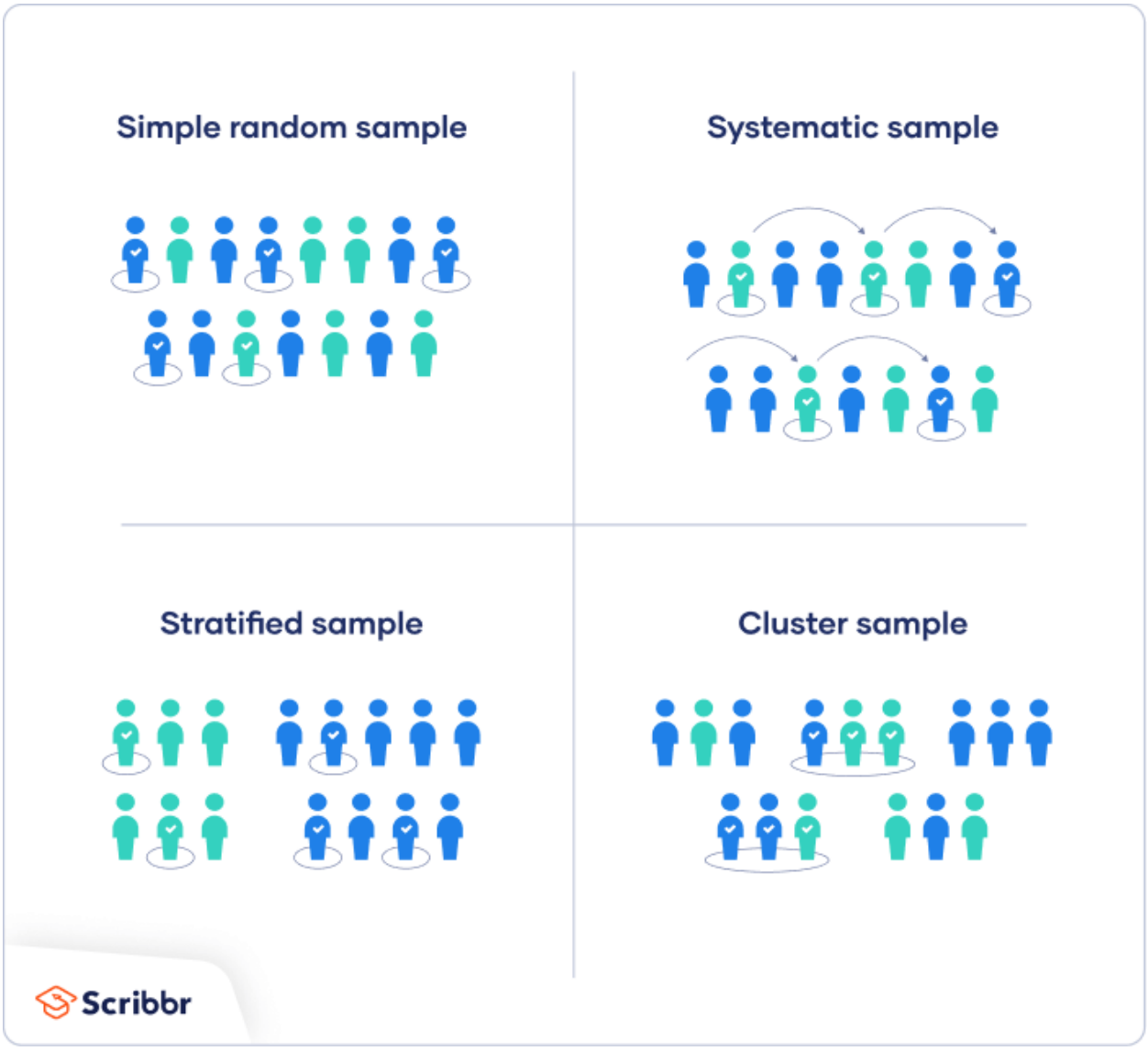
- **Probability sampling** involves random selection, allowing you to make strong statistical inferences about the whole group.
- **Non-probability sampling** involves non-random selection based on convenience or other criteria, allowing you to easily collect data.

1) Probability sampling method

Probability sampling means that every member of the population has a chance of being selected. It is mainly used in quantitative research. If you want to produce results that are representative of

the whole population, probability sampling techniques are the most valid choice. There are four main types of probability sample.

- Simple random sampling
- Systematic sampling
- Stratified sampling
- Cluster sampling



Simple random sampling

In a simple random sample, every member of the population has an equal chance of being selected. Your sampling frame should include the whole population. To conduct this type of sampling, you can use tools like random number generators or other techniques that are based entirely on chance. Sampling can be done either sampling with replacement or sampling without replacement.

Sampling without replacement

If the selected items or units are not replaced before the next draw, such a sampling is called without replacement.

Sampling with replacement

If the selected items or units are replaced before the next draw, such a sampling is called sampling with replacement

Following is the procedure for selecting simple random sampling,

1) Lottery Method

Suppose that we must select a random sample of size n from a finite population of size *N*. First assign numbers 1 to N to all the N units of the population. Then write numbers 1 to *N* on different identical slips or cards so that a card is not distinguishable from another. They are folded and mixed up in a drum or a box or a container. A blindfold selection is made. The selected card may be replaced or may not be replaced before the next draw. The Required numbers of slips are selected for the desired sample size under any one of these methods forms a simple random sample. The selection of items thus depends on chance.

2) Table of Random numbers

The easiest way to select a sample randomly is to use random number tables. There are several tables of random numbers. Some of them are

- Kendall and Smith random number table.
- Tippet's random number table.
- Fishers and Yates random number table.

Method of using Random Number table.

In a given finite population, the method of selecting a random sample from a random number table is given below:

The table contains different numbers consisting of 0, 1, 2, ,9 and possesses the essential characteristics which ensures random sampling. A part of the table is given as Appendix. Referring to this page, selection of a random sample is explained by taking a finite population of size 100 units and selecting a sample of 10 units. The steps to be followed are listed as under.

- A list of all 100 units in the finite population is prepared and each unit is assigned a serial number ranging from 00 to 99.
- Any number in the table is chosen at random and it is the starting point for selecting the sampling units.
- From the starting point we can make a move on to the next number either vertically, horizontally, or diagonally.
- Since our numbered population consists of two digits from 00 to 99, we confine ourselves to reading only two digit numbers without omitting any number that comes forward.
- As we proceed, random numbers read from the table which are above 99 are ignored and those numbers which are less than are equal to 99 are recorded. This process is continued until we reach 10 such random numbers.
- If sampling without replacement scheme is followed, a random number once recorded appears again, the same is omitted and we move on to the next number.
- The 10 such selected random numbers are compared with the labeled (numbered)
- population units and the corresponding units are selected to get a simple random sample of size 10.

Merits

- It is more representative of the universe.
- It is free from personal bias and prejudices.
- The method is simple to use.
- It is to assess sampling error in the method.

Demerits

- If the units are widely dispersed, the sample becomes unrepresentative.
- The method is not applicable when the units are heterogeneous in nature.

Systematic sampling

Systematic sampling is similar to simple random sampling, but it is usually slightly easier to conduct. Every member of the population is listed with a number, but instead of randomly generating numbers, individuals are chosen at regular intervals.

In systematic sampling, the population units are numbered from 1 to N in ascending order. A sampling interval, denoted by k, is determined as k=N/n, where n denotes the required sample size. Then n–1 such sampling intervals each consisting of k units will be formed. A number is selected at random from the first sampling interval. Let it be number i where $i \leq k$. This number is the random starting point for the whole selection of the sample. The unit corresponding to i is the first unit in the sample. The subsequent sampling units are the units in the following positions:

$i, k + i, 2k + i, 3k + i, \ldots, nk$

The layout for systematic sampling								
Sampling interval 1	1	2	3	4	...	*i* - Random start	...	k
Sampling interval 2	*k*+1	*k*+2	*k*+3	*k*+4	...	*k*+*i*	...	2k
Sampling interval 3	2*k*+1	2*k*+2	2*k*+3	2*k*+4	...	2*k*+*i*	...	3k
Sampling interval 4	3*k*+1	3*k*+2	3*k*+3	3*k*+4	...	3*k*+*i*	...	4*k*
	...	...	...	...	...	...	...	...
Sampling interval *n*	(*n*-1)*k*+1	(*n*-1)*k*+2	(*n*-1)*k*+3	(*n*-1)*k*+4	...	(*n*-1)*k*+*i*	...	*nk*

Thus, with selection of the first unit, the whole sample is selected automatically. As the first unit could have been any of the k units, the technique will generate k systematic samples with equal probability. If N is not an integral multiple of n, then sizes of a few possible systematic samples may vary by one unit.

Example: Suppose a systematic random sample of size $n = 10$ is needed from a population of size $N = 200$, the sampling interval k=N/n=200/10=20.

The first sampling interval consists of numbers 1 to 20. If the randomly selected number (random starter) is 7, the systematic sample will consist of units corresponding to positions 7, 27,47,67,87,107,127,14 7,167,187.

Merits

- This method is simple and convenient.
- Less time consuming.
- It can be used in infinite population.

Demerits

- Since it is a quasi-random sampling, the sample may not be a representative sample.

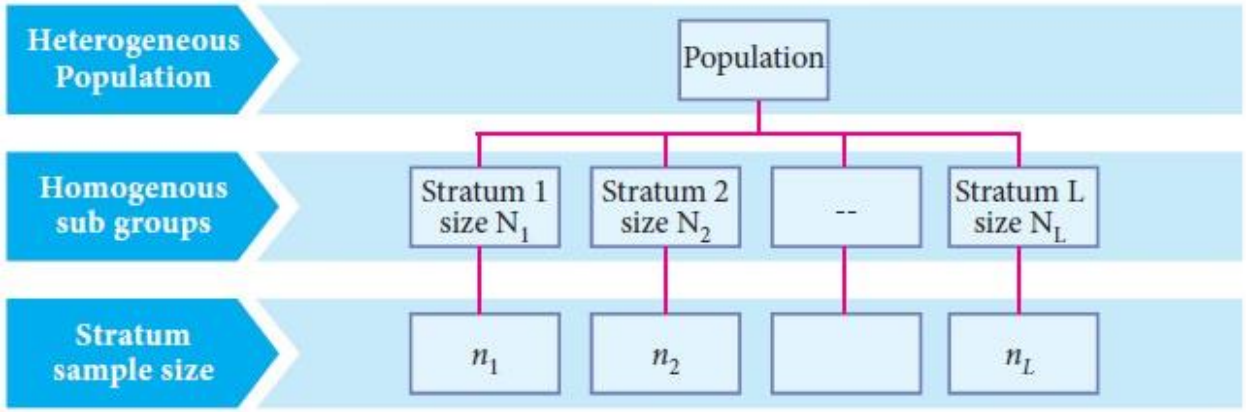
Applications:

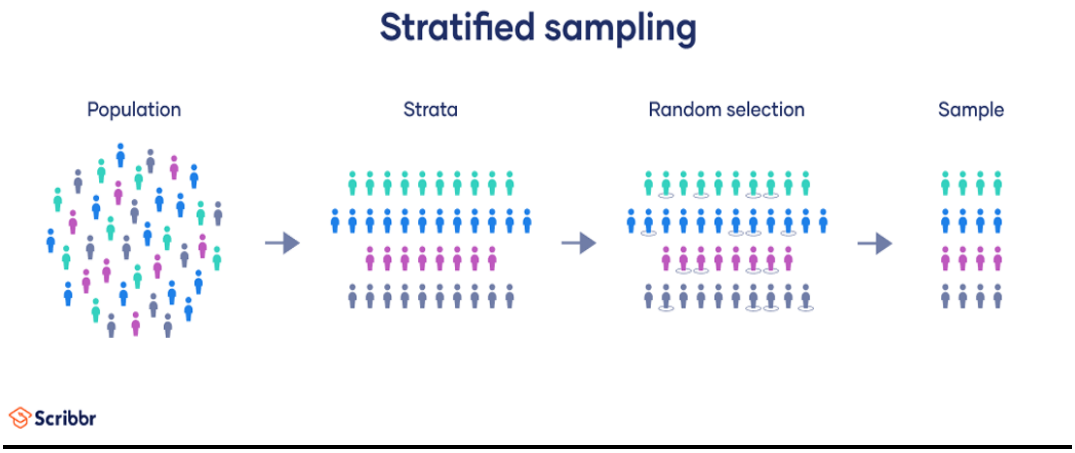
- Systematic sampling is preferably used when the information is to be collected from trees in a highway, houses in blocks, etc.,
- This method is often used in industry, where an item is selected for testing from a production line (say, every fifteenth item in the order of production) to ensure that equipment is working satisfactorily.
- This technique could also be used in a sample survey for interviewing people. A market researcher might select every 10th person who enters a particular store, after selecting a person at random as a random start.

Stratified sampling

In a **stratified sample**, researchers divide a population into homogeneous subpopulations called *strata* based on specific characteristics (e.g., race, gender identity, location, etc.). Every member of the population studied should be in exactly one stratum. Then you use random or systematic sampling to select a sample from each subgroup. Researchers rely on stratified sampling when a population's characteristics are diverse, and they want to ensure that every characteristic is properly represented in the sample.

In Stratified random sampling, the heterogeneous population of size N units is sub-divided into L homogeneous non overlapping sub populations called Strata, the i^{th} stratum having N_i units (i =1, 2, 3…, L) such that $N_1+N_2+\ldots+N_L = N$. Sample size being n_i from i^{th} stratum (i=1, 2…, L) is independently taken by simple random sampling in such way that $n_1+n_2+\ldots+n_L = n$. A sample obtained using this procedure is called a stratified random sample. If the stratum size large, that stratum will get more representation in the sample. If the stratum size small, that stratum will get less representation in the sample. The sample size for the i^{th} stratum can be determined using the formula $n_i = (n/N) * N_i$.





Example: To study about the introduction of NEET exam, the opinions are collected from 3 schools. The strength of the schools is 2000, 2500 and 4000. It is fixed that the sample size is 170. Calculate the sample size for each school?

Solution

Here $N = 2000 + 2500 + 4000 = 8500$ and $n = 170$ then $n_1 = n_2 = n_3 = ?$

$N_1 = 2000$, $N_2 = 2500$, $N_3 = 4000$

$n_1 = (n/N) \times N_1 = (170 / 8500) \times 2000 = 40$

$n_2 = (n/N) \times N_2 = (170 / 8500) \times 2500 = 50$

$n_3 = (n/N) \times N_3 = (170 / 8500) \times 4000 = 80$

Therefore 40 students from school 1, 50 students from school 2 and 80 students from school 3, are to be selected using SRS to obtain the required stratified random sample

Merits

- It provides a chance to study of all the sub-populations separately.
- An optimum size of the sample can be determined with a given cost, precision and reliability.
- It is a more precise sample.
- Biases reduced and greater precise.

Demerits

- There is a possibility of faulty stratification and hence the accuracy may be lost.
- Proportionate stratification requires accurate information on the proportion of population in each stratum.
- This method requires more time and cost.

Cluster sampling

In cluster sampling, researchers divide a population into smaller groups known as clusters. Cluster sampling also involves dividing the population into subgroups, but each subgroup should have similar characteristics to the whole sample. Instead of sampling individuals from each subgroup, you randomly select entire subgroups.

2) non-Probability sampling method

Non-probability sampling is defined as a sampling technique in which the researcher selects samples based on the subjective judgment of the researcher rather than random selection. It is a less stringent method. This sampling method depends heavily on the expertise of the researchers. It is carried out by observation, and researchers use it widely for qualitative research.

Characteristics

- Samples are selected by the researcher himself and not randomly
- These methods are not scientific in nature
- These method involve personal bias
- These are more useful in case of pilot studies
- These methods are prone to sampling error

Types of non-probability sampling

- Convenience sampling
- Judgment sampling
- Quota sampling
- Snowball sampling

Convenience sampling:

Convenience sampling is a non-probability sampling technique where samples are selected from the population only because they are conveniently available to the researcher. Researchers choose these samples just because they are easy to recruit, and the researcher did not consider selecting a sample that represents the entire population. Mostly used in case of public opinion surveys or pilot studies. For example, asking people coming out of the theatres, about the movie.

Judgement sampling:

The judgmental sampling method, researchers select the samples based purely on the researcher's knowledge and credibility. This method involves personal bias. For example, a research wants to conduct a study on the families living in a particular village. Then he selects some of the families as a part of his sample on his own.

Quota sampling:

This is another non-probability sampling method. In this method, the population is divided into different groups and the interviewer assign quotas to each group. The selection of individuals from each group is based on the judgment of the interviewer. This type of sampling is called quota sampling. Specified sizes of number of certain types of peoples are included in the sample.

Merits:

- The selection of the sample in this method is quick, easy and cheaper
- May control sample characteristics
- More chance of representative.

Demerits

- Selection bias.
- The sample is not a true representative and statistical properties cannot be applied.

Snowball sampling:

Snowball sampling helps researchers find a sample when they are difficult to locate. Researchers use this technique when the sample size is small and not easily available. This sampling system works like the referral program. Once the researchers find suitable subjects, he asks them for assistance to seek similar subjects to form a considerably good size sample. This method is

prevalent in those cases where the respondents are difficult to find because of the illegal activity or due to the social embarrassment. For example, in case of research to find drug users or AIDS suffers or cheaters who leaked a paper.

Merits:

- Appropriate for small, specialized population.
- Useful in studies involving respondents rare to find.

Demerits

- It takes more time
- Most likely not representative
- Members of the population, who are little known, disliked or whose opinions conflict with the respondents, have low probability of being included.

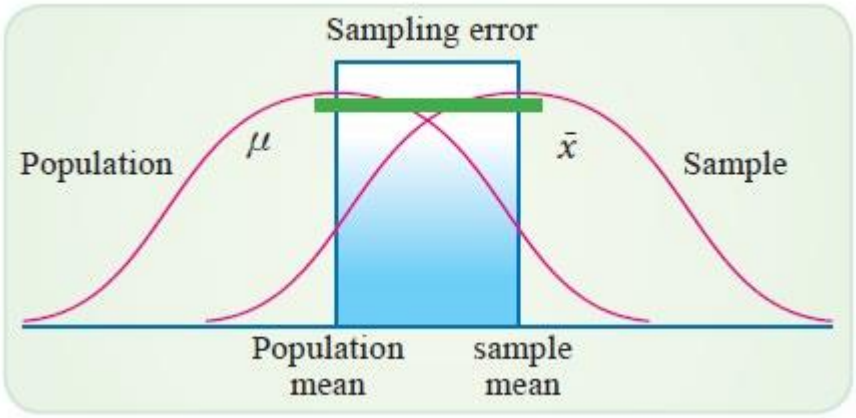
Difference between probability sampling and non-probability sampling

Probability sampling	Non-probability sampling
The sample is selected at random.	Sample selection based on the subjective judgment of the researcher.
Everyone in the population has an equal chance of getting selected.	Not everyone has an equal chance to participate.
Used when sampling bias has to be reduced.	The researcher does not consider sampling bias.
Useful when the population is diverse.	Useful when the population has similar traits.
Used to create an accurate sample.	The sample does not accurately represent the population.
Finding the right respondents is not easy.	Finding respondents is easy.

Sampling and Non-sampling errors

Sampling error

The purpose of sample is to study the population characteristics. The sample size is not equal to population size except in the case complete enumeration. Therefore, the statistical measurements like mean of the sample and mean of the population differ.



If $\bar{x}$ is the sample mean and μ is the population mean of the characteristic X, then the sampling error is $\bar{x}$ - μ. The sampling error may be positive or negative or zero.

Non-Sampling Errors

The non-sampling errors arise due to various causes right from the beginning stage when the survey is planned and designed to the final stage where the data are processed and analyzed. Non sampling errors are more serious than the sampling errors because a sampling error can be minimized by taking a large sample. It is difficult to minimize non sampling errors, even if a large sample is taken. The main sources of non-sampling errors are now described.

(i) Non-response:

The errors due to non-response may occur due to omission or lapse on the part of the interviewer, or the refusal on the part of the respondents to questions or because of the non-availability of the individuals during the period of survey.

(ii) Errors in measurement:

The measuring device may be biased or inaccurate. The respondent may not know the correct answer and may give imprecise answers. Common examples are questions on age, income, and events that happened in the past. The interviewer may also fail to record the responses correctly. Errors in measurement include errors in coding, editing, and tabulation.

Coverage errors:

The coverage errors are classified as ‘under coverage errors’ and ‘over coverage errors.

Under-coverage errors occur in the following situations:

- The selected unit in the sampling frame is not interviewed by the investigator.
- The selected unit is incorrectly classified as ineligible for surveys
- The unit is omitted or skipped by the interviewer.

Similarly, **over-coverage** occurs under the following situations:

- The sampling frame covers ineligible units.
- The frame may contain the same unit more than once.

The errors cannot be ignored since the cumulative effect of these errors affect the objectives of the survey.

Organizing a sample survey

The above said things provide a comprehensive idea about collection of data. However, when one decides to collect data through sampling the following steps are to be followed.

Stage I: Developing a sample plan

Definite sequence of steps the interviewer ought to go in order to draw and ultimately arrive at the final sample.

- Define the relevant population.
- Obtain a population list, if possible: may only be some type of a sampling frame.
- Fix the sample size
- Choose the appropriate sampling.
- Draw the sample.
- Assess the validity of the sample.
- Resample if necessary.

Stage II: Pilot survey or Pilot Study

It is a guiding survey, usually on a small scale, carried out before the main survey. The information received by pilot survey is utilized in improving the efficiency of the large scale main survey. It helps in:

- Estimating the cost of the regular survey
- Correcting the questionnaire of the survey
- Training the field workers.
- Removing the faults of the field organization.
- Deciding about the other details of the survey.

Stage III. Dealing with Non- respondents

- Procedures will have to be devised to deal with those who do not give information.