Sampling Frame

The sampling frame (also known as the “sample frame” or “survey frame”) is indeed the actual collection of units. A sample has now been taken from this. A basic random sample gives all units in it an equal probability of being drawn and appearing in the sample.

A complete list or collection from which our sample participants will be drawn in a predetermined manner. The list will be organized in some way. That is, each member of a population will have an individual identity and a contact mechanism. This allows you to categorize and code known information about segmentation features.

Collecting the sample indicates that we have a supply or list of all the individuals of the target population from which to take a sample, as well as a process for selecting the sample. Any resource that has the information needed to reach every individual in the targeted group qualifies as a source.

Characteristics of a Good Sampling Frame

Be assertive when selecting lists! Make sure the sample frame is large enough for our requirements. A decent sample frame for research on living conditions, for example, might include:

Everyone is in the target demographic.

Exclude everyone who isn’t part of the target group.

A file containing factual information that may be used to reach specific people.

Other considerations:

Each member has a unique identification. This might be a short number code (e.g., from 1 to 3000).

Make sure the frame doesn’t have any duplicates.

The list should be well organized. Sort them alphabetically for better access

Information should be up to date. This might need to be examined regularly (e.g., for address or contact number changes).

Examples of the Sampling Frame

The issue is that studying every individual in a population is not always practical or practicable.

Suppose we might be curious to learn about the opinions of Nepalese bankers about vehicle ownership, for example. Gathering data from every bank in the Nepal would be too timeconsuming and expensive. You can investigate a sample of the population in situations like these.

The process of picking a sample should be intentional, and you can utilize various sampling strategies based on the research's aim.

It would help if we first constructed a sampling frame, which would be a list of all the units in the population of interest before we can choose a sample. Our study findings can only benefit the population identified by the sample frame.

Conclusion

A sampling frame is a researcher's list or device to specify the population of interest. A basic random sample gives all units an equal probability of being drawn and appearing in the sample.

People, organizations, and existing records might all be considered units. It is critical to be as detailed as possible when describing the population.

Issues of choosing appropriate sampling technique(s) while selecting samples

Choosing a sampling strategy is an essential step in the capture phase of the data journey and will ensure that, data is reliable and reflects the characteristics of target group. In this blog, we'll take step by step through the process by outlining the ways in which primary data is collected using an example in which a survey on characteristics (tax, education levels, etc) is collected on residents in five towns. The towns are of different sizes and have a total of 3,200 households. These 3,200 households make up the target population for survey.

Step one: Define sample and target population

At times, the survey may require covering the entire target population, as is the case in mapping or population studies. That's usually referred to as a census survey. However, target populations are generally large and expensive to survey. In our example, it may not be feasible to visit all 3,200 households of the five towns. Instead, we want to choose a smaller sample that would be representative of the population and reflect its characteristics.

A survey that is done on a smaller number of the target population is referred to as a sample survey. We can infer our findings for the entire population based on this representative sample.

Step two: Define sample size

The first step in sampling exercise will be decided on an appropriate sample size. There are no strict rules for selecting a sample size. We can make a decision based on the objectives of the project, time available, budget, and the necessary degree of precision.

In order to select the appropriate sample size, we will need to determine the degree of accuracy that we want to achieve. For this, we'll need to establish the confidence interval and confidence level of our sample.

The confidence interval, also called the margin of error, is a plus or minus figure. It is the range within which the likelihood of a response occurs. The most commonly used confidence interval is +/- 5. If we wish to increase the precision level of our data, we would further reduce the error margin or confidence interval to a +/- 2. For example, if our survey question is “does the household pay tax?” and 65% of our sampled households say “yes,” then using a confidence interval of +/- 5, we can state with confidence that if we are asked the question to all 3,200 households, between 60% (i.e. 65-5) and 70% (i.e. 65+5) would have also responded “yes.”

The confidence level tells how sure we want to be and is expressed as a percentage. It represents how often the responses from our selected sample reflect the responses of the total population. Thus, a 95% confidence level means we can be 95% certain. The lower the confidence level, the less certain we will be.

Most surveys use the 95% confidence level and a +/- 5 confidence interval. When we put the confidence level and the confidence interval together, we can say that we are 95% sure that, if we had surveyed all (3,200) households, between 60% and 70% of the households of the target population would have answered “yes,” to the question “does the household pay tax?”.

The size of sample may be determined using any standard sample size calculator. Using a standard sample size calculator (as can be seen in table one below) for our example of 3,200 households in five towns, we can examine the difference in sample sizes based on different confidence levels and intervals.

Option A

If we decide on a 5% confidence interval and want to achieve a 95% confidence level, the sample size will be 345 households.

Option B

If we wish to have higher accuracy and increase the confidence level to 99%, the recommended sample size would be 551.

The quality of our findings are likely to only be marginally better than with option A or B, as the rate of improvement of accuracy gradually diminishes with the increase in sample size. The size of sample should therefore be decided by the objectives of the study and resources available.

Step three: Define sampling technique

Once we’ve chosen the sample size for survey, we need to define which sampling technique to select sample from the target population. The sampling technique that’s right depends on the nature and objectives of project. Sampling techniques can be broadly divided into two types: random sampling and non-random sampling.

Random sampling

As the name suggests, random sampling literally means selection of the sample randomly from a population, without any specific conditions. This may be done by selecting the sample from a list, such as a directory, or physically at the location of the survey. If we want to ensure that a particular household does not get selected more than once, you can remove it from the list. This type of sampling is called simple random sampling without replacement. If you choose not to remove duplicate households from the list, you would do a simple random sampling with replacement.

Systematic sampling is the most commonly used method of random sampling, whereby you divide the total population by the sample size and arrive at a figure which becomes the sampling interval for selection. For example, if you need to choose 20 samples from a total population of 100, your sampling interval would be five. Systematic sampling works best when the population is homogeneous, i.e. most people share the same characteristics. In our example, the sampling interval would be nine (3200/345= 9 for a 95% confidence level and 5% confidence interval). Thus we will select every ninth household in a town.

However, populations are generally mixed and heterogeneous. To ensure sufficient inclusion of all categories of the population, we need to identify the different strata or characteristics and their actual representation (i.e. proportion) in the population. In such cases, we can use the stratified random sampling technique, whereby we first calculate the proportion of each strata within the population and then select the sample in the same proportion, randomly or systematically, from all the strata.

If we take our earlier example of five towns, to calculate a stratified random sample, you will need to calculate the proportion of each town within the sample size of 345 as shown in table two below. Column three gives the proportion of each town of the total population (3,200). In column four, the sample size (345) is proportionately divided across the five towns. For example, town three, which is 25% of the total population, will select 86 households with a sampling interval of nine (i.e. 800/86) in the same manner as was done for systematic sampling.

Table 2: Calculate stratified random sample

Location	Population size	Proportion (%) of population	Stratified sample size
Town 1	1200	38%	129
Town 2	900	28%	97
Town 3	800	25%	86
Town 4	180	6%	19
Town 5	120	4%	13
Total	*3200*		*345*

Non-random sampling

In non-random sampling, the sample selection follows a particular set of conditions and is generally used in studies where the sample needs to be collected based on a specific characteristic of the population. For example, you may need to select only households which own a car, or have children less than six years of age. For this, you would consciously select only the 345 or 551 households that have those characteristics. Also termed purposive or subjective sampling, non-random sampling methods include convenience, judgment, quota and snowball sampling.

Step four: Minimize sampling error

It’s normal to make mistakes during sample selection. Our efforts should always be to reduce the sampling error and make the chosen sample as representative of the population as possible. The robustness of sample depends on how we minimize the sampling error. The extents of errors during sampling vary according to the technique or method you choose for sample selection.

For samples selected randomly from a target population, the results are generally prefixed with the +/- sampling error, which is the degree to which the sample differs from the population. If our study requires to know the extent of sampling error that is acceptable for the survey, you can select a random sampling technique. In random sampling, you will be able to regulate the survey design to arrive at an acceptable level of error. In a non-random sample selection, the sampling error remains unknown.

Thus, when your sample survey needs to infer the proportion of a certain characteristics of the target population, you can select a random sampling method. But if you want to know the perceptions of residents regarding taxation laws or the school curriculum, you would want to capture as many perceptions as possible, and therefore select a non-random method in situations where sampling errors or sampling for proportionality are not of concern. Non-random sampling techniques can be very useful in situations when you need to reach a targeted sample with specified characteristics very quickly.

If you don’t have a sampling strategy in place, you may collect data which is biased or not representative, rendering your data invalid.