

UNIT III

DATA COLLECTION METHOD

Primary Data

A **primary data** source is an original **data** source, that is, one in which the **data** are collected firsthand by the researcher for a specific research purpose or project. **Primary data** can be collected in a number of ways.

Secondary Data

Secondary data is the data that has already been collected through primary sources and made readily available for researchers to use for their own research. It is a type of data that has already been collected in the past.

SECONDARY DATA TYPES

Survey Data: Definition

Survey data is defined as the resultant data that is collected from a sample of respondents that took a survey. This data is comprehensive information gathered from a target audience about a specific topic to conduct research. There are many methods used for survey data collection and statistical analysis.

Various mediums are used to collect feedback and opinions from the desired sample of individuals. While conducting survey research, researchers prefer multiple sources to gather data such as online surveys, telephonic surveys, face-to-face surveys, etc. The medium of collecting survey data decides the sample of people that are to be reached out to, to reach the requisite number of survey responses.

Factors of collecting survey data such as how the interviewer will contact the respondent (online or offline), how the information is communicated to the respondents etc. decide the effectiveness of gathered data.

Observation Data

Observation is a technique that involves systematically selecting, watching, listening, reading, touching, and recording behavior and characteristics of living beings, objects, or phenomena.

The researchers, adopting this method, attempt to understand behavior and societies by getting to know the persons involved and their values, rituals, symbols, beliefs, and emotions.

Advantages of the Observation Method

Directness

The main advantage of observation is its directness. We can collect data at the time they occur. The observer does not have to ask people about their behavior and reports from others. He or she can simply watch as individuals act and speak. While the survey respondents may have a hazy or lapse memory about events that occurred in the distant past, the observer is studying events as they occur.

Natural environment

Whereas other data collection techniques introduce artificiality into the research environment, data collected in an observation study describe the observed phenomena as they occur in their natural settings.

Observation is neither as restrictive nor as artificial as either the survey or the experiment.

Longitudinal analysis

Since the observation is possible to be conducted in a natural setting, the observer can conduct his or her study over a much longer period than with either the survey or experiment.

Non-verbal behavior

Observation is decidedly superior to survey research, experimentation, or document study for collecting data on nonverbal behavior. Some studies focus on individuals who are unable to give verbal reports or to articulate themselves meaningfully.

Disadvantages of the Observation Method

Lack of control

Despite the advantage as achieved from the natural environment, the observation study, however, has little control over extraneous variables that may affect the data.

The presence of a stranger (the observer) and the error involved in human observation and the recording of data, which may remain out of control of the observer, are likely to bias the observations to a great extent.

Difficulties in quantification

Measurement in observational studies generally takes the form of observer's unquantified perceptions rather than the quantitative measures often used in the survey and experimental studies.

Smallness in sample size

Because observational studies are generally conducted in-depth, with data that are often subjective and difficult to quantify, the sample size is usually kept at a minimum.

Also, the in-depth nature of the observation studies generally requires that they are conducted over an extended period, then the survey method or experiments. This feature tends to limit the size of the sample.

No opportunity to learn past

In an observational study, there is no way to know the past. It is also difficult to gather information on such topics as intentions, opinions, attitudes, or preferences.

This technique can generate either quantitative or qualitative data but tends to be used more for small-scale exploratory studies than for large-scale quantitative studies. This is because it usually requires

- Relatively highly skilled observers and analysts
- Prolonged periods of observation
- High cost per unit of observation.

General Accuracy of Data Collected

Data quality management is the key to any company's success. It is important that one understands the effect of enterprise data accuracy problems and takes steps to solve them. Some of the key ways in which data accuracy rates can be improved are listed here -

1. Inaccurate Data Sources

Companies should identify the right data sources, both internally and externally, to improve the quality of incoming data. Incorrect data may result from migration of data from one database to another, presence of incorrect values, or even time-bound data changes. It is important to identify the cause for incorrect data and try to fix it.

2. Set Data Quality Goals

It is important that firms set realistic goals towards improving overall data quality. The top management needs to understand the basic problems which are plaguing their data accuracy and set realistic goals for the data entry specialist team. They should be examined based on efficient data capturing, data entry, and effective coding.

3. Avoid Overloading

A manager must ensure that the data entry team is not under pressure to deliver expected results from the get go. If data entry specialists are overloaded with work, they can become tired and that can result in data entry errors. If there is more work it is a good option to delegate some of the work to others in the data entry team, or span it over the course of multiple weeks.

4. Review the Data

Reviewing is an efficient way to check the correctness of the data. Companies must incorporate an efficient way to review and double check the data entered. It is always beneficial to hire a team of quality assurance professionals who can review the data and help in reducing the data errors to a large extent.

5. Automate Error Reports

Making use of advanced software is always a plus point for any company. Generating automated error reports is a common practice among leading companies today. This is especially beneficial when you are entering the same kind of data for a project.

6. Adopt Accuracy Standards

Companies must adopt highly robust data entry quality standards such as matching, geo-coding, data monitoring, data profiling, linking, etc. This ensures that the data entered conforms to pre-defined data standards which in turn help in improving the data quality.

7. Have a Good Work Environment

Having a good and healthy work environment helps the employees make lesser mistakes and therefore has a direct impact on data accuracy. Companies need to provide a healthy work environment to their data entry professionals which help in retaining their focus.

Questionnaire Method

- A **questionnaire** is defined as a document containing questions and other types of items designed to solicit information appropriate for analysis.
- The questionnaire may be regarded as a form of an interview on paper.
- Procedure for the construction of a questionnaire follows a pattern similar to that of the interview schedule.
- However, because the questionnaire is impersonal it is all the more important to take care of its construction.

- Since there is no interviewer to explain ambiguities or to check misunderstandings, the questionnaire must be especially clear in its working.
- The variety of possible answers to each question must be anticipated more fully than for an interview.

Structured questionnaires

Structured questionnaires include pre-coded questions with well-defined skipping patterns to follow the sequence of questions. Most of the quantitative data collection operations use structured questionnaires. Fewer discrepancies, easy to administer consistency in answers and easy for the data management are advantages of such structured questionnaires.

Unstructured questionnaires

Unstructured questionnaires include open-ended and vague opinion-type questions. Maybe questions are not in the format of interrogative sentences and the moderator or the enumerator has to elaborate the sense of the question. Focus group discussions use such a questionnaire.

Telephone Interview

Telephone interview is a data collection method when the interviewer communicates with the respondent on the **telephone** in accordance with the prepared questionnaire. Usually, standardized questionnaires with closed-ended questions are recommended for this kind of questioning.

Personal Interview

A personal interview survey, also called as a face-to-face survey, is a survey method that is utilized when a specific target population is involved. The purpose of conducting a personal interview survey is to explore the responses of the people to gather more and deeper information.

Personal interview surveys are used to probe the answers of the respondents and at the same time, to observe the behavior of the respondents, either individually or as a group. The personal interview method is preferred by researchers for a couple of advantages.

Questionnaire construction methods

QUESTIONNAIRE

Definitions

- ✓ A set of predetermined questions for all respondents that serves as a primary research instrument in survey research.
- ✓ Used to collect factual information
- ✓ Consist of a form containing a series of questions
- ✓ A questionnaire is a series of questions asked to individuals to obtain statistically useful information about a given topic. When properly constructed and responsibly administered, questionnaires become a vital instrument by which statements can be made about specific groups or people or entire populations

Characteristics of a good questionnaire

1. Should be concerned with specific and relevant topic
2. Should be short
3. Directions and wording should be simple and clear
4. Questions should be objective
5. Embarrassing questions, presuming questions and hypothetical questions should be avoided
6. Should be presented in a good order
7. Should be attractive, neatly printed and clearly arranged

Functions of questionnaire

DESCRIPTION

The questionnaire provides description about age, sex, marital status, occupation, income, political affiliation religious affiliation, etc.

UNIT IV

Sampling

Sampling is a process used in statistical analysis in which a predetermined number of observations are taken from a larger population. The methodology used to **sample** from a larger population depends on the type of analysis being performed, but it may include simple random **sampling** or systematic **sampling**.

TYPES OF SAMPLING

Probability Sampling Methods

1. Simple random sampling

In this case each individual is chosen entirely by chance and each member of the population has an equal chance, or probability, of being selected. One way of obtaining a random sample is to give each individual in a population a number, and then use a table of random numbers to decide which individuals to include.¹ For example, if you have a sampling frame of 1000 individuals, labelled 0 to 999, use groups of three digits from the random number table to pick your sample. So, if the first three numbers from the random number table were 094, select the individual labelled “94”, and so on.

As with all probability sampling methods, simple random sampling allows the sampling error to be calculated and reduces selection bias. A specific advantage is that it is the most straightforward method of probability sampling. A disadvantage of simple random sampling is that you may not select enough individuals with your characteristic of interest, especially if that characteristic is uncommon. It may also be difficult to define a complete sampling frame and inconvenient to contact them, especially if different forms of contact are required (email, phone, post) and your sample units are scattered over a wide geographical area.

2. Systematic sampling

Individuals are selected at regular intervals from the sampling frame. The intervals are chosen to ensure an adequate sample size. If you need a sample size n from a population of size x , you should select every x/n^{th} individual for the sample. For example, if you wanted a sample size of 100 from a population of 1000, select every $1000/100 = 10^{\text{th}}$ member of the sampling frame.

Systematic sampling is often more convenient than simple random sampling, and it is easy to administer. However, it may also lead to bias, for example if there are underlying patterns in the order of the individuals in the sampling frame, such that the sampling technique coincides with the periodicity of the underlying pattern. As a hypothetical example, if a group of students were being sampled to gain their opinions on college facilities, but the Student Record Department's central list of all students was arranged such that the sex of students alternated between male and female, choosing an even interval (e.g. every 20th student) would result in a sample of all males or all females. Whilst in this example the bias is obvious and should be easily corrected, this may not always be the case.

3. Stratified sampling

In this method, the population is first divided into subgroups (or strata) who all share a similar characteristic. It is used when we might reasonably expect the measurement of interest to vary between the different subgroups, and we want to ensure representation from all the subgroups. For example, in a study of stroke outcomes, we may stratify the population by sex, to ensure equal representation of men and women. The study sample is then obtained by taking equal sample sizes from each stratum. In stratified sampling, it may also be appropriate to choose non-equal sample sizes from each stratum. For example, in a study of the health outcomes of nursing staff in a county, if there are three hospitals each with different numbers of nursing staff (hospital A has 500 nurses, hospital B has 1000 and hospital C has 2000), then it would be appropriate to choose the sample numbers from each hospital *proportionally* (e.g. 10 from hospital A, 20 from hospital B and 40 from hospital C). This ensures a more realistic and accurate estimation of the health outcomes of nurses across the county, whereas simple random sampling would over-represent nurses from hospitals A and B. The fact that the sample was stratified should be taken into account at the analysis stage.

Stratified sampling improves the accuracy and representativeness of the results by reducing sampling bias. However, it requires knowledge of the appropriate characteristics of the sampling frame (the details of which are not always available), and it can be difficult to decide which characteristic(s) to stratify by.

4. Clustered sampling

In a clustered sample, subgroups of the population are used as the sampling unit, rather than individuals. The population is divided into subgroups, known as clusters, which are randomly selected to be included in the study. Clusters are usually already defined, for example individual GP practices or towns could be identified as clusters. In single-stage cluster sampling, all members of the chosen clusters are then included in the study. In two-stage cluster sampling, a selection of individuals from each cluster is then randomly selected for inclusion. Clustering should be taken into account in the analysis. The General Household survey, which is undertaken annually in England, is a good example of a (one-stage) cluster sample. All members of the selected households (clusters) are included in the survey.¹

Cluster sampling can be more efficient than simple random sampling, especially where a study takes place over a wide geographical region. For instance, it is easier to contact lots of individuals in a few GP practices than a few individuals in many different GP practices. Disadvantages include an increased risk of bias, if the chosen clusters are not representative of the population, resulting in an increased sampling error.

Non-Probability Sampling Methods

1. Convenience sampling

Convenience sampling is perhaps the easiest method of sampling, because participants are selected based on availability and willingness to take part. Useful results can be obtained, but the results are prone to significant bias, because those who volunteer to take part may be different from those who choose not to (volunteer bias), and the sample may not be representative of other characteristics, such as age or sex. Note: volunteer bias is a risk of all non-probability sampling methods.

2. Quota sampling

This method of sampling is often used by market researchers. Interviewers are given a quota of subjects of a specified type to attempt to recruit. For example, an interviewer might be told to go out and select 20 adult men, 20 adult women, 10 teenage girls and 10 teenage boys so that they could interview them about their television viewing. Ideally the quotas chosen would proportionally represent the characteristics of the underlying population.

Whilst this has the advantage of being relatively straightforward and potentially representative, the chosen sample may not be representative of other characteristics that weren't considered (a consequence of the non-random nature of sampling).²

3. Judgement (or Purposive) Sampling

Also known as selective, or subjective, sampling, this technique relies on the judgement of the researcher when choosing who to ask to participate. Researchers may implicitly thus choose a “representative” sample to suit their needs, or specifically approach individuals with certain characteristics. This approach is often used by the media when canvassing the public for opinions and in qualitative research.

Judgement sampling has the advantage of being time-and cost-effective to perform whilst resulting in a range of responses (particularly useful in qualitative research). However, in addition to volunteer bias, it is also prone to errors of judgement by the researcher and the findings, whilst being potentially broad, will not necessarily be representative.

4. Snowball sampling

This method is commonly used in social sciences when investigating hard-to-reach groups. Existing subjects are asked to nominate further subjects known to them, so the sample increases in size like a rolling snowball. For example, when carrying out a survey of risk behaviours amongst intravenous drug users, participants may be asked to nominate other users to be interviewed.

Snowball sampling can be effective when a sampling frame is difficult to identify. However, by selecting friends and acquaintances of subjects already investigated, there is a significant risk of selection bias (choosing a large number of people with similar characteristics or views to the initial individual identified).

Bias in sampling

There are five important potential sources of bias that should be considered when selecting a sample, irrespective of the method used. Sampling bias may be introduced when:¹

1. Any pre-agreed sampling rules are deviated from
2. People in hard-to-reach groups are omitted

3. Selected individuals are replaced with others, for example if they are difficult to contact
4. There are low response rates
5. An out-of-date list is used as the sample frame (for example, if it excludes people who have recently moved to an area)

Sampling problems

Problems due to sampling bias Sampling bias is problematic because it is possible that a statistic computed of the sample is systematically erroneous. Sampling bias can lead to a systematic over- or under-estimation of the corresponding parameter in the population.

SMPLE ERROR

Sampling error is a statistical error that occurs when an analyst does not select a sample that represents the entire population of data. The results found in the sample thus do not represent the results that would be obtained from the entire population.

Choosing a Sample Design

For most research problems, there is not just one possible research design, but a range of possibilities to choose from. The choices you make depend on your priorities in the research, and often involve some tradeoffs – a research design that is strong in one area might be weaker in another.