

Chapter Four: Data Collection Methods and Analysis

4.1 Sources and Types of data

Data are the raw materials of research. Although we are surrounded by data, it is not so straightforward to collect the correct data for our purposes. It needs a plan of action that identifies and uses the most effective and appropriate methods of data collection. The task of data collection begins after a research problem has been defined and research design/plan chalked out. While deciding about the method of data collection to be used for the study, the researcher should keep in mind two types of data and data sources viz., primary and secondary. The *primary data* are those which are collected from primary sources which are fresh and collected for the first time, and thus happen to be original in character. The *secondary data*, on the other hand, are those which have already been collected by someone else and which have already been passed through the statistical process. Hence, secondary data are collected from secondary sources. The researcher would have to decide which sort of data he would be using (thus collecting) for his study and accordingly he will have to select one or the other method of data collection. The methods of collecting primary and secondary data differ since primary data are to be originally collected, while in case of secondary data the nature of data collection work is merely that of compilation of secondary sources.

4.2 Data collection Instruments/Methods

There are various instruments of data collection. As such the researcher must judiciously select the method/methods or instruments for his own study, keeping in view the following factors:

1. **Nature, scope and object of enquiry:** This constitutes the most important factor affecting the choice of a particular method or instrument. The method selected should be such that it suits the type of enquiry that is to be conducted by the researcher. This factor is also important in deciding whether the data already available (secondary data) are to be used or the data not yet available (primary data) are to be collected.
2. **Availability of funds:** Availability of funds for the research project determines to a large extent the method to be used for the collection of data. When funds at the disposal of the researcher are very limited, he will have to select a comparatively cheaper method which may not be as efficient and effective as some other costly method. Finance, in fact, is a big constraint in practice and the researcher has to act within this limitation.
3. **Time factor:** Availability of time has also to be taken into account in deciding a particular method of data collection. Some methods take relatively more time, whereas with others the data can be collected in a comparatively shorter duration. The time at the disposal of the researcher, thus, affects the selection of the method by which the data are to be collected.
4. **Precision required:** Precision required is yet another important factor to be considered at the time of selecting the method of collection of data.

But one must always remember that each method of data collection has its uses and none is superior in all situations. For instance, telephone interview method may be considered appropriate (assuming telephone population) if funds are restricted, time is also restricted and the data is to be collected in respect of few items with or without a certain degree of precision. In case funds permit and more information is desired, personal interview method may be said to be relatively better. In case time is ample, funds are limited and much information is to be gathered with no precision, then mail-questionnaire method can be regarded more reasonable. When funds are ample, time is also ample and much information with no precision is to be collected, then either personal interview or the mail-questionnaire or the joint use of these two methods may be taken as an appropriate method of collecting data. Where a wide geographic area is to be covered, the use of mail-questionnaires supplemented by personal interviews will yield more reliable results per rupee spent than either method alone. The secondary data may be used in case the researcher finds them reliable, adequate and appropriate for his research. While studying motivating influences in market researches or studying people's attitudes in psychological/social surveys, we can resort to the use of one or more of the projective techniques stated earlier. Such techniques are of immense value in case the reason is obtainable from the respondent who knows the reason but does not want to admit it or the reason relates to some underlying psychological attitude and the respondent is not aware of it. But when the respondent knows the reason and can tell the same if asked, than a non-projective questionnaire, using direct questions, may yield satisfactory results even in case of attitude surveys. Since projective techniques are as yet in an early stage of development and with the validity of many of them remaining an open question, it is usually considered better to rely on the straight forward statistical methods with only supplementary use of projective techniques. Nevertheless, in pre-testing and in searching for hypotheses they can be highly valuable. Thus, the most desirable approach with regard to the selection of the method depends on the nature of the particular problem and on the time and resources (money and personnel) available along with the desired degree of accuracy. But, over and above all this, much depends upon the ability and experience of the researcher.

Collection of Primary Data

We collect primary data during the course of doing experiments in an experimental research but in case we do research of the descriptive type and perform surveys, whether sample surveys or census surveys, then we can obtain primary data either through observation or through direct communication with respondents in one form or another or through personal interviews. This means that there are several methods of collecting primary data, particularly in surveys and descriptive researches. Important ones are: (i) observation method, (ii) interview method, (iii) through questionnaires, (iv) through schedules, and (v) Focus group discussion, (vi) Key informant Interview, etc.

1. Observation Method

The observation method is the most commonly used method especially in studies relating to behavioural sciences. In a way we all observe things around us, but this sort of observation is not scientific observation. Observation becomes a scientific tool and the method of data collection for the researcher, when it serves a formulated research purpose, is systematically planned and recorded and is subjected to checks and controls on validity and reliability. Under the observation method, the information is sought by way of investigator's own direct observation without asking from the respondent.

The main advantage of this method is that subjective bias is eliminated, if observation is done accurately. Secondly, the information obtained under this method relates to what is currently happening; it is not complicated by either the past behaviour or future intentions or attitudes. Thirdly, this method is independent of respondents' willingness to respond and as such is relatively less demanding of active cooperation on the part of respondents as happens to be the case in the interview or the questionnaire method. This method is particularly suitable in studies which deal with subjects (i.e., respondents) who are not capable of giving verbal reports of their feelings for one reason or the other. In addition to this, observation is important for geographers to identify all geographical settings and happenings.

However, observation method has various limitations. Firstly, it is an expensive method. Secondly, the information provided by this method is very limited. Thirdly, sometimes unforeseen factors may interfere with the observational task. At times, the fact that some people are rarely accessible to direct observation creates obstacle for this method to collect data effectively. While using this method, the researcher should keep in mind things like: What should be observed? How the observations should be recorded? Or how the accuracy of observation can be ensured? In case the observation is characterised by a careful definition of the units to be observed, the style of recording the observed information, standardised conditions of observation and the selection of pertinent data of observation, then the observation is called as *structured observation*. But when observation is to take place without these characteristics to be thought of in advance, the same is termed as *unstructured observation*. Structured observation is considered appropriate in descriptive studies, whereas in an exploratory study the observational procedure is most likely to be relatively unstructured. We often talk about participant and non-participant types of observation in the context of studies, particularly of social sciences. This distinction depends upon the observer's sharing or not sharing the life of the group he is observing. If the observer observes by making himself, more or less, a member of the group he is observing so that he can experience what the members of the group experience, the observation is called as the *participant observation*. But when the observer observes as a detached emissary without any attempt on his part to experience through participation what others feel, the observation of this type is often termed as *non-participant observation*. (When the observer is observing in such a manner that his presence may be unknown to the people he is observing, such an observation is described as *disguised observation*.) There are several merits of the participant type of observation: (i) The researcher is enabled to record the natural behaviour of the group. (ii) The researcher can even gather information which could not easily be obtained if he

observes in a disinterested fashion. (iii) The researcher can even verify the truth of statements made by informants in the context of a questionnaire or a schedule. But there are also certain demerits of this type of observation viz., the observer may lose the objectivity to the extent he participates emotionally; the problem of observation-control is not solved; and it may narrow-down the researcher's range of experience. Sometimes we talk of *controlled* and *uncontrolled observation*. If the observation takes place in the natural setting, it may be termed as uncontrolled observation, but when observation takes place according to definite pre-arranged plans, involving experimental procedure, the same is then termed controlled observation. In non-controlled observation, no attempt is made to use precision instruments. The major aim of this type of observation is to get a spontaneous picture of an area, life and persons. It has a tendency to supply naturalness and completeness of behaviour, allowing sufficient time for observing it. But in controlled observation, we use mechanical (or precision) instruments as aids to accuracy and standardisation. Such observation has a tendency to supply formalised data upon which generalisations can be built with some degree of assurance. The main pitfall of non-controlled observation is that of subjective interpretation. There is also the danger of having the feeling that we know more about the observed phenomena than we actually do. Generally, controlled observation takes place in various experiments that are carried out in a laboratory or under controlled conditions, whereas uncontrolled observation is resorted to in case of exploratory researches.

2. Interview Method

The interview method of collecting data involves presentation of oral-verbal stimuli and reply in terms of oral-verbal responses. This method can be used through personal interviews and, if possible, through telephone interviews.

3. Collection of Data through Questionnaires

This method of data collection is quite popular, particularly in case of big enquiries. It is being adopted by private individuals, research workers, private and public organisations and even by governments. In this method a questionnaire is sent (usually by post or individuals) to the persons concerned with a request to answer the questions and return the questionnaire. A questionnaire consists of a number of questions printed or typed in a definite order on a form or set of forms. The questionnaire is mailed to respondents who are expected to read and understand the questions and write down the reply in the space meant for the purpose in the questionnaire itself. The respondents have to answer the questions on their own.

The method of collecting data by mailing the questionnaires to respondents is most extensively employed in various economic and business surveys. The merits claimed on behalf of this method are as follows:

1. There is low cost even when the universe is large and is widely spread geographically.
2. It is free from the bias of the interviewer; answers are in respondents' own words.
3. Respondents have adequate time to give well thought out answers.
4. Respondents, who are not easily approachable, can also be reached conveniently.
- 5 Large samples can be made use of and thus the results can be made more dependable and reliable.

The main demerits or limitations of this system can also be listed here:

1. Low rate of return of the duly filled in questionnaires; bias due to no-response is often indeterminate.
2. It can be used only when respondents are educated and cooperating.
3. The control over questionnaire may be lost once it is sent.
4. There is inbuilt inflexibility because of the difficulty of amending the approach once questionnaires have been despatched.
5. There is also the possibility of ambiguous replies or omission of replies altogether to certain questions; interpretation of omissions is difficult.
6. It is difficult to know whether willing respondents are truly representative.
7. This method is likely to be the slowest of all.

Before using this method, it is always advisable to conduct 'pilot study' (Pilot Survey) for testing the questionnaires. In a big enquiry the significance of pilot survey is felt very much. Pilot survey is in fact the replica and rehearsal of the main survey. Such a survey, being conducted by experts, brings to the light the weaknesses (if any) of the questionnaires and also of the survey techniques. From the experience gained in this way, improvement can be effected.

Main aspects of a Questionnaire

Quite often questionnaire is considered as the heart of a survey operation. Hence it should be very carefully constructed. If it is not properly set up, then the survey is bound to fail. This fact requires us to study the main aspects of a questionnaire viz., the general form, question sequence and question formulation and wording. Researcher should note the following with regard to these three main aspects of a questionnaire:

1. *General form:* So far as the general form of a questionnaire is concerned, it can either be structured or unstructured questionnaire. Structured questionnaires are those questionnaires in which there are definite, concrete and pre-determined questions. The questions are presented with exactly the same wording and in the same order to all respondents. Resort is taken to this sort of standardisation to ensure that all respondents reply to the same set of questions. The form of the question may be either closed (i.e., of the type 'yes' or 'no') or open (i.e., inviting free response) but should be stated in advance and not constructed during questioning. Structured questionnaires may also have fixed alternative questions in which responses of the informants are limited to the stated alternatives. Thus a highly structured questionnaire is one in which all questions and answers are specified and comments in the respondent's own words are held to the minimum. When these characteristics are not present in a questionnaire, it can be termed as unstructured or non-structured questionnaire. More specifically, we can say that in an unstructured questionnaire, the interviewer is provided with a general guide on the type of information to be obtained, but the exact question formulation is largely his own responsibility and the replies are to be taken down in the respondent's own words to the extent possible; in some situations tape recorders may be used to achieve this goal.

Structured questionnaires are simple to administer and relatively inexpensive to analyse. The provision of alternative replies, at times, helps to understand the meaning of the question

clearly. But such questionnaires have limitations too. For instance, wide range of data and that too in respondent's own words cannot be obtained with structured questionnaires. They are usually considered inappropriate in investigations where the aim happens to be to probe for attitudes and reasons for certain actions or feelings. They are equally not suitable when a problem is being first explored and working hypotheses sought. In such situations, unstructured questionnaires may be used effectively. Then on the basis of the results obtained in pre test (testing before final use) operations from the use of unstructured questionnaires, one can construct a structured questionnaire for use in the main study.

2. *Question sequence*: In order to make the questionnaire effective and to ensure quality to the replies received, a researcher should pay attention to the question-sequence in preparing the questionnaire. A proper sequence of questions reduces considerably the chances of individual questions being misunderstood. The question-sequence must be clear and smoothly-moving, meaning thereby that the relation of one question to another should be readily apparent to the respondent, with questions that are easiest to answer being put in the beginning. The first few questions are particularly important because they are likely to influence the attitude of the respondent and in seeking his desired cooperation. The opening questions should be such as to arouse human interest. The following type of questions should generally be avoided as opening questions in a questionnaire:

1. Questions that put too great a strain on the memory or intellect of the respondent;
2. Questions of a personal character;
3. Questions related to personal wealth, etc.

Following the opening questions, we should have questions that are really vital to the research problem and a connecting thread should run through successive questions. Ideally, the question sequence should conform to the respondent's way of thinking. Knowing what information is desired, the researcher can rearrange the order of the questions (this is possible in case of unstructured questionnaire) to fit the discussion in each particular case. But in a structured questionnaire the best that can be done is to determine the question-sequence with the help of a Pilot Survey which is likely to produce good rapport with most respondents. Relatively difficult questions must be relegated towards the end so that even if the respondent decides not to answer such questions, considerable information would have already been obtained. Thus, question-sequence should usually go from the general to the more specific and the researcher must always remember that the answer to a given question is a function not only of the question itself, but of all previous questions as well. For instance, if one question deals with the price usually paid for coffee and the next with reason for preferring that particular brand, the answer to this latter question may be couched largely in terms of price differences.

3. *Question formulation and wording*: With regard to this aspect of questionnaire, the researcher should note that each question must be very clear for any sort of misunderstanding can do irreparable harm to a survey. Question should also be impartial in order not to give a biased picture of the true state of affairs. Questions should be constructed with a view to their forming a logical part of a well thought out tabulation plan. In general, all questions should meet the following standards—(a) should be easily understood; (b) should be simple i.e.,

should convey only one thought at a time; (c) should be concrete and should conform as much as possible to the respondent's way of thinking. (For instance, instead of asking, "How many razor blades do you use annually?" The more realistic question would be to ask, "How many razor blades did you use last week?")

Concerning the form of questions, we can talk about two principal forms, viz., multiple choice question and the open-end question. In the former the respondent selects one of the alternative possible answers put to him, whereas in the latter he has to supply the answer in his own words. The question with only two possible answers (usually 'Yes' or 'No') can be taken as a special case of the multiple choice question, or can be named as a 'closed question.' There are some advantages and disadvantages of each possible form of question. Multiple choice or closed questions have the advantages of easy handling, simple to answer, quick and relatively inexpensive to analyse. They are most amenable to statistical analysis. Sometimes, the provision of alternative replies helps to make clear the meaning of the question. But the main drawback of fixed alternative questions is that of "putting answers in people's mouths" i.e., they may force a statement of opinion on an issue about which the respondent does not in fact have any opinion. They are not appropriate when the issue under consideration happens to be a complex one and also when the interest of the researcher is in the exploration of a process. In such situations, open-ended questions which are designed to permit a free response from the respondent rather than one limited to certain stated alternatives are considered appropriate. Such questions give the respondent considerable latitude in phrasing a reply. Getting the replies in respondent's own words is, thus, the major advantage of open-ended questions. But one should not forget that, from an analytical point of view, open-ended questions are more difficult to handle, raising problems of interpretation, comparability and interviewer bias. In practice, one rarely comes across a case when one questionnaire relies on one form of questions alone. The various forms complement each other. As such questions of different forms are included in one single questionnaire. For instance, multiple-choice questions constitute the basis of a structured questionnaire, particularly in a mail survey. But even there, various open-ended questions are generally inserted to provide a more complete picture of the respondent's feelings and attitudes.

Researcher must pay proper attention to the wordings of questions since reliable and meaningful returns depend on it to a large extent. Since words are likely to affect responses, they should be properly chosen. Simple words, which are familiar to all respondents, should be employed. Words with ambiguous meanings must be avoided. Similarly, danger words, catch-words or words with emotional connotations should be avoided. Caution must also be exercised in the use of phrases which reflect upon the prestige of the respondent. Question wording, in no case, should bias the answer. In fact, question wording and formulation is an art and can only be learnt by practice.

4. Collection of Data Through Schedules

This method of data collection is very much like the collection of data through questionnaire, with little difference which lies in the fact that schedules (proforma containing a set of questions) are being filled in by the enumerators who are specially appointed for the purpose. These enumerators along with schedules go to respondents, put to them the questions from the proforma in the order the questions are listed and record the replies in the space meant for the same in the proforma. In certain situations, schedules may be handed over to respondents and enumerators may help them in recording their answers to various questions in the said schedules. Enumerators explain the aims and objects of the investigation and also remove the difficulties which any respondent may feel in understanding the implications of a particular question or the definition or concept of difficult terms.

This method requires the selection of enumerators for filling up schedules or assisting respondents to fill up schedules and as such enumerators should be very carefully selected. The enumerators should be trained to perform their job well and the nature and scope of the investigation should be explained to them thoroughly so that they may well understand the implications of different questions put in the schedule. Enumerators should be intelligent and must possess the capacity of cross examination in order to find out the truth. Above all, they should be honest, sincere, and hardworking and should have patience and perseverance.

This method of data collection is very useful in extensive enquiries and can lead to fairly reliable results. It is, however, very expensive and is usually adopted in investigations conducted by governmental agencies or by some big organisations. Population census all over the world is conducted through this method.

5. Depth interviews

Depth interviews are those interviews that are designed to discover underlying motives and desires and are often used in motivational research. Such interviews are held to explore needs, desires and feelings of respondents. In other words, they aim to elicit unconscious as also other types of material relating especially to personality dynamics and motivations. As such, depth interviews require great skill on the part of the interviewer and at the same time involve considerable time. Unless the researcher has specialised training, depth interviewing should not be attempted. Depth interview may be projective in nature or it may be a non-projective interview. The difference lies in the nature of the questions asked. Indirect questions on seemingly irrelevant subjects provide information that can be related to the informant's behaviour or attitude towards the subject under study. Thus, for instance, the informant may be asked on his frequency of air travel and he might again be asked at a later stage to narrate his opinion concerning the feelings of relatives of some other man who gets killed in an airplane accident. Reluctance to fly can then be related to replies to questions of the latter nature. If the depth interview involves questions of such type, the same may be treated as projective depth interview. But in order to be useful, depth interviews do not necessarily have to be projective in nature; even non-projective depth interviews can reveal important aspects of psycho-social situation for understanding the attitudes of people.

Collection of Secondary Data

Secondary data means data that are already available i.e., they refer to the data which have already been collected and analysed by someone else. When the researcher utilises secondary data, then he has to look into various sources from where he can obtain them. In this case he is certainly not confronted with the problems that are usually associated with the collection of original data. Secondary data may either be published data or unpublished data. Usually published data are available in: (a) various publications of the central, state or local governments; (b) various publications of foreign governments or of international bodies and their subsidiary organisations; (c) technical and trade journals; (d) books, magazines and newspapers; (e) reports and publications of various associations connected with business and industry, banks, stock exchanges, etc.; (f) reports prepared by research scholars, universities, economists, etc. in different fields; and (g) public records and statistics, historical documents, and other sources of published information. The sources of unpublished data are many; they may be found in diaries, letters, unpublished biographies and autobiographies and also may be available with scholars and research workers, trade associations, labour bureaus and other public/private individuals and organisations.

Researcher must be very careful in using secondary data. He must make a minute scrutiny because it is just possible that the secondary data may be unsuitable or may be inadequate in the context of the problem which the researcher wants to study.

By way of caution, the researcher, before using secondary data, must see that they possess following characteristics:

1. Reliability of data: The reliability can be tested by finding out such things about the said data:

- (a) Who collected the data?
- (b) What were the sources of data?
- (c) Were they collected by using proper methods?
- (d) At what time were they collected?
- (e) Was there any bias of the compiler?
- (f) What level of accuracy was desired? Was it achieved?

2. Suitability of data: The data that are suitable for one enquiry may not necessarily be found suitable in another enquiry. Hence, if the available data are found to be unsuitable, they should not be used by the researcher. In this context, the researcher must very carefully scrutinise the definition of various terms and units of collection used at the time of collecting the data from the primary source originally. Similarly, the object, scope and nature of the original enquiry must also be studied. If the researcher finds differences in these, the data will remain unsuitable for the present enquiry and should not be used.

3. Adequacy of data: If the level of accuracy achieved in data is found inadequate for the purpose of the present enquiry, they will be considered as inadequate and should not be used by the researcher.

The data will also be considered inadequate, if they are related to an area which may be either narrower or wider than the area of the present enquiry. From all this we can say that it is very risky to use the already available data. The already available data should be used by the

researcher only when he finds them reliable, suitable and adequate. But he should not blindly discard the use of such data if they are readily available from authentic sources and are also suitable and adequate for in that case it will not be economical to spend time and energy in field surveys for collecting information. At times, there may be wealth of usable information in the already available data which must be used by an intelligent researcher but with due precaution.

4.4 Guide to construct data collection instruments

Guidelines for Constructing Questionnaire/Schedule

The researcher must pay attention to the following points in constructing an appropriate and effective questionnaire or a schedule:

1. The researcher must keep in view the problem he is to study for it provides the starting point for developing the Questionnaire/Schedule. He must be clear about the various aspects of his research problem to be dealt with in the course of the research.
2. Appropriate form of questions depends on the nature of information sought, the sampled respondents and the kind of analysis intended. The researcher must decide whether to use closed or open-ended question. Questions should be simple and must be constructed with a view to their forming a logical part of a well thought out tabulation plan. The units of enumeration should also be defined precisely so that they can ensure accurate and full information.
3. Rough draft of the Questionnaire/Schedule is prepared, giving due thought to the appropriate sequence of putting questions. Questionnaires or schedules previously drafted (if available) may as well be looked into at this stage.
4. Researcher must invariably re-examine, and in case of need may revise the rough draft for a better one. Technical defects must be minutely scrutinised and removed.
5. Pilot study should be undertaken for pre-testing the questionnaire. The questionnaire may be edited in the light of the results of the pilot study.
6. Questionnaire must contain simple but straight forward directions for the respondents so that they may not feel any difficulty in answering the questions.

Guidelines for Successful Interviewing

Interviewing is an art and one learns it by experience. However, the following points may be kept in view by an interviewer for eliciting the desired information:

1. Interviewer must plan in advance and should fully know the problem under consideration. He must choose a suitable time and place so that the interviewee may be at ease during the interview period. For this purpose some knowledge of the daily routine of the interviewee is essential.
2. Interviewer's approach must be friendly and informal. Initially friendly greetings in accordance with the cultural pattern of the interviewee should be exchanged and then the purpose of the interview should be explained.
3. All possible effort should be made to establish proper rapport with the interviewee; people are motivated to communicate when the atmosphere is favourable.

4. Interviewer must know that ability to listen with understanding, respect and curiosity is the gateway to communication, and hence must act accordingly during the interview. For all this, the interviewer must be intelligent and self-disciplined.
5. There should be a free-flowing interview and the questions must be well-phrased in order to have full cooperation of the interviewee. But the interviewer must control the course of the interview in accordance with the objective of the study.
6. In case of big enquiries, where the task of collecting information is to be accomplished by several interviewers, there should be an interview guide.

4.5 Ethical Practices in Geographic Research

Introducing yourself is the first ethical practice in open research. If you are relying on someone else to find participants for you, it is important that you make sure that that person knows who you are and what you are doing and that this information is then passed on to everyone else. A useful way to do this is to produce a leaflet which can be given to anyone who might be thinking about taking part in your research. This leaflet should contain the following information:

- Details of who you are (student and course or employee and position).
- Details of the organisation for which you work or at which you study.
- Information about who has commissioned/funded the research, if relevant.
- Information about your project – subject and purpose.
- Details about what will happen to the results.
- Information about the personal benefits to be gained by taking part in the project. This section is optional, but it is found that it helps to show that people will gain personally in some way by taking part in the research. This acts as an incentive.

You might offer further information about something in which they are interested, or you might offer them a copy of the final report. Some consumer research companies offer entry into a prize draw or vouchers for local shops and restaurants.

Treating participants with respect is the second major ethical practice in research. As a researcher you must remember that the research process intrudes on people's lives. Some of the people who take part in your research may be vulnerable because of their age, social status or position of powerlessness. If participants are young, you need to make sure a parent or guardian is present. If participants are ill or reaching old age you might need to use a proxy and care should be taken to make sure that you do not affect the relationship between the proxy and the participant. Some people may find participation a rewarding process, whereas others will not. Your research should not give rise to false hopes or cause unnecessary anxiety. You must try to minimise the disruption to people's lives and if someone has found it an upsetting experience you should find out why and try to ensure that the same situation does not occur again. As a researcher you will encounter awkward situations, but good preparation and self-awareness will help to reduce these. If they do happen, you should not dwell too long on the negative side – reflect, analyse, learn by your mistakes and move on.

Anonymity and confidentiality is the other very crucial ethical practice. You must do your best to ensure anonymity and confidentiality. However, information given by research participants in confidence does not enjoy legal privilege. This means that the information may be liable to subpoena by a court. If you're dealing with very sensitive information which you know could be called upon by a court of law, you will need to inform your participants that you would be obliged to hand over the information.

The fourth major ethical practices are recognising overt and covert research. Overt research means that it is open, out in the public and that everyone knows who you are and what you are doing. On the other hand, covert research means that you are doing it under cover, that no one knows you are a researcher or what you are doing. According to individuals opinion covert research should be kept to a minimum – there are enough journalists and television personalities doing this kind of undercover, sensationalist work. In the past researchers have justified their covert work by saying that it has been the only way to find out what goes on in a particular organisation that would not otherwise let a researcher enter. Such work has been carried out within religious cults and warring gangs of young people. However, this type of research can have serious implications for the personal safety of the researcher and the people with whom she comes into contact. It can also give research a bad name-other people may read about the work and become suspicious about taking part in future projects.

It is believed that researchers should be open and honest about who they are and what they are doing it overt research. People can then make an informed choice about whether they take part in a project. It is their prerogative to refuse – nobody should be forced, bullied or cajoled into doing something they don't want to do. If people are forced to take part in a research project, perhaps by their boss or someone else in a position of authority, you will soon find out. They will not be willing to participate and may cause problems for you by offering false or useless information, or by disrupting the data collection process. Who can blame them? Wouldn't you do the same if you were forced to do something you didn't want to do? This means that not only should you be open and honest about who you are and what you're doing, but so should those who open the gates for you, especially those who are in a position of authority.

4.6 Methods of data analysis

The data, after collection, has to be processed and analyzed in accordance with the outline laid down for the purpose at the time of developing the research plan. Technically speaking, processing implies editing, coding, classification and tabulation of collected data so that they are amenable to analysis. The term analysis refers to the computation of certain measures along with searching for patterns of relationship that exist among data-groups. Thus, “in the process of analysis, relationships or differences supporting or conflicting with original or new hypotheses should be subjected to statistical tests of significance to determine with what validity data can be said to indicate any conclusions”.

The way of identifying relationships or differences supporting or conflicting with original or new hypotheses can be through qualitative or quantitative analysis.

Qualitative Data Analysis

Qualitative data analysis is defined as "working with data, organizing it, breaking it into manageable units, synthesizing it, searching for patterns, discovering what is important and what is to be learned, and deciding what you will tell others“. Qualitative researchers tend to use inductive analysis of data, meaning that the critical themes emerge out of the data.

Qualitative analysis requires some creativity, for the challenge is to place the raw data into logical, meaningful categories; to examine them in a holistic fashion; and to find a way to communicate this interpretation to others. For qualitative data, the researcher might analyze as the research progresses, continually refining and reorganizing in light of the emerging results. The format for analysis could be a transcript from an:

- ✓ interview or focus group,
- ✓ a series of written answers on an open-ended questionnaire, or
- ✓ field notes or
- ✓ Memos written by the researcher.

Types of qualitative Analysis

1. Thematic analysis: Data is analyzed by theme and it is highly inductive, that is, the themes emerge from the data and are not imposed upon it by the researcher. The data collection and analysis take place simultaneously.
2. Comparative Analysis: Data from different people is compared and contrasted and the process continues until the researcher is satisfied that no new issues are arising. Comparative and thematic analyses are often used in the same project, with the researcher moving backwards and forwards between transcripts, memos, notes and the research literature.
3. Content analysis: In content analysis, the analysis being left until the data has been collected; and the most common method of doing this is to code by content. This type of analysis can be used for open-ended questions which have been added to

questionnaires in large quantitative surveys, thus enabling the researcher to quantify the answers.

4. Discourse analysis/ conversational analysis

These methods look at patterns of speech, such as how people talk about a particular subject, what metaphors they use, how they take turns in conversation, and so on.

These analysts see speech as a performance; it performs an action rather than describes a specific state of affairs or specific state of mind. Much of this analysis is intuitive and reflective, but it may also involve some form of counting, such as counting instances of turn-taking and their influence on the conversation and the way in which people speak to others.

Processing the data and Stages of data analysis

It is possible to undertake the mechanical process using computing software which can save you a lot of time, although it may stop you becoming really familiar with the data.

There are many dedicated qualitative analysis programs of various kinds available to social researchers that can be used for a variety of different tasks. For example, software could locate particular words or phrases; make lists of words and put them into alphabetical order; insert key words or comments; count occurrences of words or phrases or attach numeric codes. Some software will retrieve text, some will analyse text and some will help to build theory. Although a computer can undertake these mechanical processes, it cannot think about, judge or interpret qualitative data.

Stages of data analysis

1. Open Coding

Analysis begins with identification of the themes emerging from the raw data, a process sometimes referred to as "open coding". During open coding, the researcher must identify and tentatively name the conceptual categories into which the phenomena observed will be grouped. The goal is to create descriptive, multi-dimensional categories which form a preliminary framework for analysis. Words, phrases or events that appear to be similar can be grouped into the same category. These categories may be gradually modified or replaced during the subsequent stages of analysis that follow.

2. Axial coding

The next stage of analysis involves re-examination of the categories identified to determine how they are linked, a complex process sometimes called "axial coding".

The discrete categories identified in open coding are compared and combined in new ways as the researcher begins to assemble the "big picture." The purpose of coding is to not only describe but, more importantly, to acquire new understanding of a phenomenon of interest. Therefore, causal events contributing to the phenomenon; descriptive details of the phenomenon itself; and the ramifications of the phenomenon under study must all be identified and explored. During axial coding the researcher is responsible for building a conceptual model and for determining whether sufficient data exists to support that interpretation.

3. Translation/ interpretation

Finally the researcher must translate the conceptual model into the story line that will be read by others. Ideally, the research report will be a rich, tightly woven account that closely approximates the reality it represents.

Quantitative Data Analysis

Quantitative analysis deals with numbers and uses mathematical operations to investigate the properties of data. The levels of measurement used in the collection of the data (i.e. nominal, ordinal, interval and ratio) are an important factor in choosing the type of analysis that is applicable, as is the number of cases involved. Statistics is the name given to this type of analysis, and is defined in this sense as: The science of collecting and analysing numerical data, especially in, or for, large quantities, and usually inferring proportions in a whole from proportions in a representative sample.

For quantitative data, the analysis can be left until the end of the data collection process, and if it is a large survey, statistical software is the easiest and most efficient method to use.

The most common package used by social scientists at this present time is SPSS for windows, which has become increasingly user-friendly over the last few years. SPSS is a software package used for conducting statistical analyses, manipulating data, and generating tables and graphs that summarize data. Statistical analyses range from basic descriptive statistics, such as averages and frequencies, to advanced inferential statistics, such as regression models, analysis of variance, and factor analysis. SPSS also contains several tools for manipulating data, including functions for recoding data and computing new variables, as well as for merging and aggregating datasets. SPSS also has a number of ways to summarize and display data in the form of tables and graphs. There are also, many software packages at the push of a key produce professional graphs, tables and pie charts which can be used in your final report, again saving a lot of time and effort.

Types of variable/ Measurement levels

Variables are key elements in research. A variable is defined as a characteristic of the participants or situation for a given study that has different values in that study. A variable must be able to vary or have different values or levels. For example, *gender* is a variable because it has two levels, female or male.

Age is a variable that has a large number of values. *Type of treatment/intervention* (or *type of curriculum*) is a variable if there is more than one treatment or a treatment and a control group. *Number of days to learn something or to recover from an ailment* are common measures of the effect of a treatment and, thus, are also variables. Similarly, *amount of*

mathematics knowledge is a variable because it can vary from none to a lot. If a concept has only one value in a particular study, it is not a variable; it is a constant. Thus, ethnic group is not a variable if all participants are European American. Gender is not a variable if all participants in a study are female.

In quantitative research, variables are **defined operationally** and are commonly divided into **independent variables** (active or attribute) and **dependent variables**.

Measurement is the assignment of numbers or symbols to the different characteristics (values) of variables according to rules. In order to understand your variables, it is important to know their level of measurement. Depending on the level of measurement of a variable, the data can mean different things. To help understand these differences, types or levels of variables have been identified. It is common and traditional to discuss four levels or scales of measurement: **nominal**, **ordinal**, **interval**, and **ratio**, which vary from the unordered (nominal) to the highest level (ratio).

- 1. Nominal:** Subjects of research are differentiated by possessing or not possessing a given characteristic, e.g., pass/fail, single/married, and divided into a number of categories but the difference between the categories is not measurable in any real sense. This is the least sophisticated level of measurement.
- 2. Ordinal:** Subjects are ranked in order from greatest to least or best to worst. Again there is no precisely measurable difference between the ranks.
- 3. Interval:** Genuinely quantitative measurement such as that of temperature is measured at the interval level of measurement. Here the difference between 10 and 11 degrees centigrade is the same as the difference between 11 and 12 degrees centigrade.
- 4. Ratio:** In a scale of measurement where the difference between points on the scale is precise (as in the measurement of height and weight,) *and the scale starts at zero* the level of measurement is referred to as ratio. Height and weight start at zero. You can not weigh less than 0.00kg and cannot be less than 0.00mm in length/height; these are ratio scales. You can however record temperatures of the weather in terms of minus x degrees centigrade and this is why the scale is interval and not ratio.

Data measured by nominal and ordinal methods will not be organized in a curve form. Nominal data tend to be in the dichotomous form of either/or (e.g. this is a cow or a sheep or neither), while ordinal data can be displayed in the form of a set of steps (e.g. the first, second and third positions on the winner's podium). For those cases where this parameter is absent, non-parametric statistics may be applicable.

Non-parametric statistics are tests that have been devised to recognize the particular characteristics of non-curve data and to take into account these singular characteristics by specialized methods. In general, these types of test are less sensitive and powerful than parametric tests; they need larger samples in order to generate the same level of significance.

Statistical tests (parametric)

There are two classes of parametric statistical tests: descriptive statistics, which quantify the characteristics of parametric numerical data, and inferential statistics, which produce predictions through inference based on the data analysed. Distinction is also made between the numbers of variables considered in relation to each other:

1. Univariate analysis – analyses the qualities of one variable at a time.
2. Bivariate analysis – considers the properties of two variables in relation to each other.
3. Multivariate analysis – looks at the relationships between more than two variables.

Univariate analysis (descriptive)

A range of properties of one variable can be examined using the following measures:

• **Frequency distribution** – usually presented as a table, this simply shows the values for each variable expressed as a number and as a percentage of the total of cases. SPSS provides a choice of display options to illustrate the measures listed above. The most basic is a summary table of descriptive statistics that gives figures for all of the measures. More graphical options, which make comparisons between variables simpler, are:

- i. Bar graph – this shows the distribution of nominal and ordinal variables. The categories of the variables are along the horizontal axis (x axis) and the values are on the vertical axis (y axis). The bars should not touch each other.
- ii. Histogram – a bar graph with the bars touching to produce a shape that reflects the distribution of a variable.
- iii. Frequency polygon (or frequency curve) – a line that connects the tops of the bars of a histogram to provide a pure shape illustrating distribution.
- iv. Pie chart – this shows the values of a variable as a section of the total cases (like slices of a pie). The percentages are also usually given.
- v. Box and whisker plot – this gives more detail of the values that lie within the various percentiles (10th, 25th, 50th, 75th and 90th). Individual values that are outside this range can be pinpointed manually if they are judged to be important.

• **Measure of central tendency** – is one number that denotes what is commonly called the ‘average’ of the values for a variable. There are several measures that can be used:

1. Arithmetic mean – this is the arithmetic average calculated by adding all the values and dividing by their number. This can be calculated for ordinal, interval and ratio variables.
2. Mode – the value that occurs most frequently. The only measure that can be used with nominal variables, as well as all the others.
3. Median – the mid-point in the distribution of values that is the mathematical middle between the highest and lowest value. It is used for ordinal, interval and ratio variables.

• **Measures of dispersion** (or variability) – all of the above measures are influenced by the nature of dispersion of the values (how values are spread out or bunched up) and the presence

of solitary extreme values. To investigate the dispersion, the following measures can be made:

1. Range – the distance between the highest and lowest value.
2. Inter quartile range – the distance between the value that has a quarter of the values less than it (first quartile or 25th percentile) and the value that has three-quarters of the values less than it (third quartile or 75th percentile).
3. Variance – the average of the squared deviations for the individual values from the mean.
4. Standard deviation – the square root of the variance.
5. Standard error – the standard deviation of the mean score.

Bivariate analysis: Bivariate analysis considers the properties of two variables in relation to each other. The relationship between two variables is of common interest in the social sciences, for example: Does social status influence academic achievement? Are boys more likely to be delinquents than girls? Does age have an effect on community involvement? There are various methods for investigating the relationships between two variables. An important aspect is the different measurement of these relationships, such as assessing the direction and degree of association, statistically termed correlation coefficients. The commonly used coefficients assume that there is a linear relationship between the two variables, either positive or negative. In reality, this is seldom achieved, but degrees of correlation can be computed – how near to a straight line the relationship is.

Scattergrams

Scattergrams are a useful type of diagram that graphically shows the relationship between two variables by plotting variable data from cases on a two-dimensional matrix. If the resulting plotted points appear in a scattered and random arrangement, then no association is indicated. If, however, they fall into a linear arrangement, a relationship can be assumed, either positive or negative. The closer the points are to a perfect line, the stronger the association. A line that is drawn to trace this notional line is called **the line of best fit** or **regression line**. This line can be used to predict one variable value on the basis of the other.

It is quite possible to get forms of relationships between variables that are not represented in a straight line, for example groupings or curved linear arrangements. The strength of the scattergrams is that these are clearly shown, thus needing some discussion and possible explanation. For these relationships, statistical tests that assume linearity should not be used.

Contingency tables

Cross-tabulation (contingency tables) is a simple way to display the relationship between variables that have only a few categories. The cells made by the rows show the relationships between each of the categories of the variables in both number of responses and percentages. In addition, the column and row totals and percentages are shown. These can be conveniently produced by SPSS from the data compiled on a matrix. Patterns of association can be detected if they occur. As an alternative, the display can be automatically presented as a bar chart. The choice of appropriate statistical methods of bivariate analysis depends on the levels of measurement used in the variables. Here are some of the most commonly used:

- i. Pearson's correlation coefficient (r) should be used for examining relationships between interval/ratio variables. The r value indicates the strength and direction of the correlation (how close the points are to a straight line). +1 indicates a perfect positive association and -1 a perfect negative association. Zero indicates a total lack of association.
- ii. Spearman's rho (ρ) should be used either when both variables are ordinal, or when one is ordinal and the other is interval/ratio.
- iii. Spearman rank correlation coefficient and Kendall's Tau are both used with ordinal data.
- iv. Phi (Φ) should be used when both variables are dichotomous (e.g. yes/no).
- v. Cramer's V is used when both variables are nominal and with positive values.

Statistical significance

As most analysis is carried out on data from only a sample of the population, the question is raised as to how likely is it that the results indicate the situation for the whole population. The process of testing statistical significance to generalize from a sample to the population as a whole is known as statistical inference.

The most common statistical tool for this is known as the **chi-square test**. This measures the degree of association or linkage between two variables by comparing the differences between the **observed values** and **expected values** if no association were present that is those that would be a result of pure chance. This is commonly referred to as the ***p-value*** (p standing for *probability*). The probability values are sometimes given in reports of quantitative research (e.g. $p = 0.03$ meaning that probability is less than 3 in 100). A common acceptable maximum *p-value* in social science research is 0.05, but if the researcher wants to be particularly cautious, a maximum of 0.01 is chosen. The value of chi-square is affected by sample size, that is, the bigger the sample, the greater the chance that it will be representative. In addition, for reliable results, the chi-squared calculations require that the minimum expected values of at least 20 per cent of the cells in the contingency table should be greater than 5.

Analysis of variance

The above tests are all designed to look for relationships between variables. Another common requirement is to look for differences between values obtained under two or more different conditions, for example a group before and after a training course, or three groups after different training courses. There are a range of tests that can be applied depending on the number of groups. For a single group, say the performance of students on a particular course compared with the mean results of all the other courses in the university, you can use:

- Chi-square as a test of 'goodness of fit'.
- One-group t-test, which compares the means of the results from the sample compared with the population mean. For two groups, for example comparing the results from the same course at two different universities, you can use:

- Two-group t-test, which compares the means of two groups. There are two types of test, one for paired scores (i.e. where the same persons provided scores under each condition) or for unpaired scores, where this is not the case.

For three or more groups, for example the performance of three different age groups in a test, it is necessary to identify the dependent and independent variables that will be tested. A simple test using SPSS is:

- ANOVA (analysis of variance) – this tests the difference between the means of results gained under different conditions. One-way analysis of variance is applicable when there is one dependent variable and one independent variable (e.g. a new study course) and no matter how many groups or tests are involved. For more complex situations, when more than one independent variable is involved and a single variable, then multiple-way or factorial ANOVA should be used.

Multivariate analysis

Multivariate analysis looks at the relationships between more than two variables.

First, let us look at the effect of a third variable in the relationship between two variables. Elaboration analysis method, devised by Paul Lazarfeld and his colleagues (1972), is a set of techniques that involves a set of steps that has been clearly formulated by Marsh (1982):

1. Establish a relationship between two variables (e.g. income and level of education).
2. Subdivide the data on the basis of the values of a third variable (e.g. men and women).
3. Review the original two-variable relationship for each of the sub groups (e.g. income and education among men, and income and education among women).
4. Compare the relationship found in each sub-group with the original relationship.

When presented in tabular form, the initial table (step 1) is called the zero order contingency table, for example one which shows a significant positive relationship between two variables. However, this may be a spurious result in that the result is actually influenced more by another variable that has not been taken into account. Therefore, a separate table (conditional table) is set up to test the influence of this variable on the two original ones (step 3 above).

If the two tables show a similar significant relationship between the two original variables, this is called replication – the original relationship remains. If neither table shows a significant relationship (zero-order correlation) between the two variables, the original relationship was either spurious, meaning that the test variable actually caused the association between the original variables, or that the test variable is an intervening variable, one that varies because of the independent variable and in turn effects the dependent variable.

If one of the conditional tables demonstrates the association but the other one does not, then it shows a limitation to the association of the original pair of variables, or provides a specification of the conditions under which association occurs.

You can continue the process of producing tables for fourth and fifth variables, but this quickly becomes unwieldy. It is also difficult to get enough data in each table to achieve significant results. There are better ways to understand the interactions between large numbers of variables and the relative strength of their influence, for example **regression** techniques such as multiple regression and logistic regression.

Multiple Regressions

Multiple regression is a technique used to measure the effects of two or more independent variables on a single dependent variable measured on interval or ratio scales, for example the effect on income due to age, education, ethnicity, area of living, and gender. Thanks to computer programs such as SPSS, the complicated mathematical calculations required for this analysis are done automatically. Note that it is assumed that there are interrelationships between the independent variables as well, and this is taken into account in the calculations. The result of multiple regression – the combined correlation of a set of independent variables with the dependent variable – is termed multiple R. The square of this, multiple R², indicates the amount of variance in the independent variable due to the simultaneous action of two or more independent variables.

Logistic regression

Logistic regression is a development of multiple regressions that has the added advantage of holding certain variables constant in order to assess the independent influence of key variables of interest. It is suitable for assessing the influence of independent variables on a dependent variable measured in a nominal scale (for example, whether students' decisions to do a masters degree were determined by a range of considerations such as cost, future job prospects, level of enjoyment of student life, amount of interest in the subject, etc. The statistic resulting is an odds ratio (e.g. a student who was interested in the subject was 2.1 times as likely to do a masters than one that was not, assuming all the other variables were held constant).

Path analysis

The detailed effect of the different interrelationships between independent variables on each other and subsequently on the dependent variable is not investigated in multiple regression analysis. Theories about the types and extent of these interrelationships between independent variables and the effect of these on the dependent variable can be tested with path analysis. It requires the researcher to make guesses about how the system of variables works, and then test if these guesses are correct.

The path coefficients for pairs of independent variables can be calculated and mapped to show how much changes in each independent variable influence the others and what effect these have on the dependent variable.

Factor analysis

Factor analysis is an exploratory technique used widely in the social sciences to build reliable, compact scales for measuring social and psychological variables. It is used to package information and for data reduction. Although based on complex mathematical calculations (SPSS will do the calculations for you), the idea behind the technique is simple. This is that if a number of variables correlate with each other, they must have something in common. This common thing is called a factor, a 'super-variable' one that encompasses other variables. This simplifies the explanation of the effects of a set of independent variables on a dependent variable. Factor analysis starts with a matrix of correlations. Large matrices containing numerous variables are notoriously difficult to interpret. Factor analysis makes this easier by identifying clusters of variables that show a high degree of correlation. These

clusters can be reduced to a factor. For example, the level of intelligence may be a factor in the exam results of a wide variety of students studying a range of subjects at different educational establishments over years of results. Factors of this type often represent latent (or unobserved) variables – abstract or theoretical constructs that are not directly observable but must be deduced from several other observable variables. Factor analysis is used to examine the relationship between the latent and observed variables.

Multi-dimensional scaling

Multi-dimensional scaling (MDS) is similar to factor analysis in that it reduces data by seeking out underlying relationships between variables. The difference is that MDS does not require metric data that is data measured on the interval or ratio scale. Much data about attitudes and cognition are based on ordinal measurement. By using graphical displays to chart the associations between sets of items (people, things, attitudes, etc.) the strength of association can be easily portrayed. The relationships between three variables can be plotted as a triangle – each point representing a variable and the distance between them representing the strength of association. Points closer together have a higher correlation than those further apart. A similar approach can be used for four variables, although the number of interrelationships (six) means that a three-dimensional display will provide a better picture of the correlation strengths. Obviously, it is impossible to increase the number of dimensions to match that of the variables, so a two-dimensional map based on a matrix is conventionally used to plot a large number of variables. The stress value can be calculated to gauge the amount of distortion required to reduce the display to two dimensions. The pattern of values distributed on the map is then inspected in order to identify any clusters or arrays that reveal patterns of association.

Cluster analysis

Cluster analysis is a descriptive tool that explores relationships between items on a matrix – which items go together in which order. It measures single link and complete link clustering based on data entered on to a dissimilarity matrix. The result of the analysis is a more closely measured grouping than that achieved visually by MDS. This method does not label the clusters.

Structural equation modelling

Unlike factor analysis, structural equation modelling (SEM) is a confirmatory tool, and has become ever more popular in the social sciences for the analysis of non-experimental data in order to test hypotheses. Its strength is that it provides the opportunity to estimate the extent of error in the model, such as the effects of measurement error. SEM goes a step further than factor analysis by enabling the researcher to test structural (regression) relationships between factors (i.e. between latent variables).

Analysis of variance

Just as ANOVA measured the differences between two variables, the program called **MANOVA** (multiple analysis of variance) enables you to do many types of analysis of variance with several nominal and interval variables together. It is particularly appropriate when the dependent variable is an interval measure and the predicting variables are nominal. It is also able to detect differences on a set of dependent variables instead of just one.

Chapter Five: The selection of Survey Mechanism

A "survey" is a systematic method for gathering information from (a sample of) entities for the purposes of constructing quantitative descriptors of the attributes of the larger population of which the entities are members. The word "systematic" is deliberate and meaningfully distinguishes surveys from other ways of gathering information. The phrase "(a sample of)" appears in the definition because sometimes surveys attempt to measure everyone in a population and sometimes just a sample.

Many scholars noted that the choice of survey medium is determined by the resources that are available. Hence, the medium of survey can be written or verbal. Written surveys require minimum resources (staff, time, and cost) and are best suited to eliciting confidential information. Minimal sampling error occurs due to the relatively low cost per survey. There are also minimal interviewer and respondent measurement errors due to the absence of direct contact. Written surveys may be distributed using either postal or electronic mail. In some cases, written surveys are distributed in person to a group of respondents to evaluate a recent event. This approach is frequently used in military survey research where after action reports are used to evaluate an exercise. Although this method provides immediate results, the involuntary nature of an in-person written survey makes this medium prone to response biases. Among the disadvantages of written surveys are their subjectivity to certain types of error. For example, written surveys are subject to coverage error where population lists are incomplete or out of date. They are also typically subject to non response error. Less educated, illiterate, and disabled people are particularly less likely to respond to written surveys. Verbal surveys include telephone and face-to-face interviews. The face-to-face interview is a particularly flexible tool that can capture verbal inflexion, gestures, and other body language. A skilled interviewer can obtain additional insights into the answers provided by observing the respondent's body language. Face-to-face interviews are useful where the true population is not known or when respondents are unable or unlikely to respond to written surveys.

Advantages of Surveys

Compared to observation or other qualitative methods, survey methods allow the collection of significant amounts of data in an economical and efficient manner, and they typically allow for much larger sample sizes. There are five advantages of using survey methods: (1) standardization, (2) ease of administration, (3) ability to tap the "unseen," (4) suitability to tabulation and statistical analysis, and (5) sensitivity to subgroup differences.

1. Standardization

Because questions are pre set and organized in a particular arrangement on a questionnaire, survey methods ensure that all respondents are asked the same questions and are exposed to the same response options for each question. Moreover, the researcher is assured that every respondent will be confronted with questions that address the complete range of information objectives driving the research project.

2. Ease of Administration

Sometimes an interviewer is used, and survey modes are easily geared to such administration. On the other hand, the respondent may fill out the questionnaire unattended. In either case,

the administration aspects are much simpler than, for instance, conducting a focus group or utilizing depth interviews. Perhaps the simplest case is a mail survey in which questionnaires are sent to prospective respondents. There is no need for tape recording, taking notes, or analysing projective or physiological data; there is not even a need to read the questions to the respondent. All the researcher needs to do is mail the questions to prospective respondents.

3. Ability to Tap the “Unseen”

The four questions of what, why, how, and who help uncover “unseen” data. For instance, we can ask a working parent to tell us how important the location of a preschool was in his or her selection of the child’s preschool. We can inquire as to how many different preschools he or she seriously considered before deciding on one, and we can easily gain an understanding of the person’s financial or work circumstances with a few questions on income, occupation, and family size. Much information is unobservable and requires direct questions.

4. Suitability to Tabulation and Statistical Analysis

The marketing researcher ultimately must interpret the patterns or common themes sometimes hidden in the raw data he or she collects. Statistical analysis, both simple and complex, is the preferred means of achieving this goal, and large cross-sectional surveys perfectly complement these procedures. Qualitative methods, in contrast, prove much more frustrating in this respect because of their necessarily small samples, need for interpretation, and general approach to answering marketing managers’ questions. Increasingly, questionnaire design software includes the ability to perform simple statistical analyses, such as tabulations of the answers to each question, as well as the ability to create color graphs summarizing these tabulations.

5. Sensitivity to Subgroup Differences

Because surveys involve large numbers of respondents, it is relatively easy to “slice” up the sample into demographic groups or other subgroups and then to compare them for market segmentation implications. In fact, the survey sample design may be drawn up to specifically include important subgroups as a means of looking at market segment differences. In any case, the large sample sizes that characterize surveys facilitate subgroup analyses and comparisons of various groups existing in the sample.

5.1 Personal interview survey

Personal interview method requires a person known as the interviewer asking questions generally in a face-to-face contact to the other person or persons. (At times the interviewee may also ask certain questions and the interviewer responds to these, but usually the interviewer initiates the interview and collects the information.) This sort of interview may be in the form of direct personal investigation or it may be indirect oral investigation. In the case of direct personal investigation the interviewer has to collect the information personally from

the sources concerned. He has to be on the spot and has to meet people from whom data have to be collected. This method is particularly suitable for intensive investigations. But in certain cases it may not be possible or worthwhile to contact directly the persons concerned or on account of the extensive scope of enquiry, the direct personal investigation technique may not be used. In such cases an indirect oral examination can be conducted under which the interviewer has to cross-examine other persons who are supposed to have knowledge about the problem under investigation and the information, obtained is recorded. Most of the commissions and committees appointed by government to carry on investigations make use of this method. The method of collecting information through personal interviews is usually carried out in a structured way. As such we call the interviews as *structured interviews*. Such interviews involve the use of a set of predetermined questions and of highly standardised techniques of recording. Thus, the interviewer in a structured interview follows a rigid procedure laid down, asking questions in a form and order prescribed. As against it, the *unstructured interviews* are characterised by a flexibility of approach to questioning. Unstructured interviews do not follow a system of pre-determined questions and standardised techniques of recording information. In a non-structured interview, the interviewer is allowed much greater freedom to ask, in case of need, supplementary questions or at times he may omit certain questions if the situation so requires. He may even change the sequence of questions. He has relatively greater freedom while recording the responses to include some aspects and exclude others. But this sort of flexibility results in lack of comparability of one interview with another and the analysis of unstructured responses becomes much more difficult and time-consuming than that of the structured responses obtained in case of structured interviews. Unstructured interviews also demand deep knowledge and greater skill on the part of the interviewer. Unstructured interview, however, happens to be the central technique of collecting information in case of exploratory or formulative research studies. But in case of descriptive studies, we quite often use the technique of structured interview because of its being more economical, providing a safe basis for generalisation and requiring relatively lesser skill on the part of the interviewer. We may as well talk about focussed interview, clinical interview and the non-directive interview.

Focussed interview is meant to focus attention on the given experience of the respondent and its effects. Under it the interviewer has the freedom to decide the manner and sequence in which the questions would be asked and has also the freedom to explore reasons and motives. The main task of the interviewer in case of a focussed interview is to confine the respondent to a discussion of issues with which he seeks conversance. Such interviews are used generally in the development of hypotheses and constitute a major type of unstructured interviews. The *clinical interview* is concerned with broad underlying feelings or motivations or with the course of individual's life experience. The method of eliciting information under it is generally left to the interviewer's discretion. In case of *non-directive interview*, the interviewer's function is simply to encourage the respondent to talk about the given topic with a bare minimum of direct questioning. The interviewer often acts as a catalyst to a comprehensive expression of the respondents' feelings and beliefs and of the frame of reference within which such feelings and beliefs take on personal significance.

Despite the variations in interview-techniques, the major advantages and weaknesses of personal interviews can be enumerated in a general way.

The chief merits of the interview method are as follows:

- i. More information and that too in greater depth can be obtained.
- ii. Interviewer by his own skill can overcome the resistance, if any, of the respondents; the interview method can be made to yield an almost perfect sample of the general population.
- iii. There is greater flexibility under this method as the opportunity to restructure questions is always there, especially in case of unstructured interviews.
- iv. Observation method can as well be applied to recording verbal answers to various questions.
- v. Personal information can as well be obtained easily under this method.
- vi. Samples can be controlled more effectively as there arises no difficulty of the missing returns; non-response generally remains very low.
- vii. The interviewer can usually control which person(s) will answer the questions. This is not possible in mailed questionnaire approach. If so desired, group discussions may also be held.
- viii. The interviewer may catch the informant off-guard and thus may secure the most spontaneous reactions than would be the case if mailed questionnaire is used.
- ix. The language of the interview can be adapted to the ability or educational level of the person interviewed and as such misinterpretations concerning questions can be avoided.
- x. The interviewer can collect supplementary information about the respondent's personal characteristics and environment which is often of great value in interpreting results.

But there are also certain weaknesses of the interview method. Among the important weaknesses are the following:

- i. It is a very expensive method, especially when large and widely spread geographical sample is taken.
- ii. There remains the possibility of the bias of interviewer as well as that of the respondent; there also remains the headache of supervision and control of interviewers.
- iii. Certain types of respondents such as important officials or executives or people in high income groups may not be easily approachable under this method and to that extent the data may prove inadequate.
- iv. This method is relatively more-time-consuming, especially when the sample is large and recalls upon the respondents are necessary.
- v. The presence of the interviewer on the spot may over-stimulate the respondent, sometimes even to the extent that he may give imaginary information just to make the interview interesting.

- vi. Under the interview method the organisation required for selecting, training and supervising the field-staff is more complex with formidable problems.
- vii. Interviewing at times may also introduce systematic errors.
- viii. Effective interview presupposes proper rapport with respondents that would facilitate free and frank responses. This is often a very difficult requirement.

Pre-requisites and basic tenets of interviewing: For successful implementation of the interview method, interviewers should be carefully selected, trained and briefed. They should be honest, sincere, hardworking, impartial and must possess the technical competence and necessary practical experience.

Occasional field checks should be made to ensure that interviewers are neither cheating, nor deviating from instructions given to them for performing their job efficiently. In addition, some provision should also be made in advance so that appropriate action may be taken if some of the selected respondents refuse to cooperate or are not available when an interviewer calls upon them.

In fact, interviewing is an art governed by certain scientific principles. Every effort should be made to create friendly atmosphere of trust and confidence, so that respondents may feel at ease while talking to and discussing with the interviewer. The interviewer must ask questions properly and intelligently and must record the responses accurately and completely. At the same time, the interviewer must answer legitimate question(s), if any, asked by the respondent and must clear any doubt that the latter has. The interviewers approach must be friendly, courteous, conversational and unbiased. The interviewer should not show surprise or disapproval of a respondent's answer but he must keep the direction of interview in his own hand, discouraging irrelevant conversation and must make all possible effort to keep the respondent on the track.

5.2 Mail survey

A mail survey is one in which the questions are mailed to prospective respondents who are asked to fill them out and return them to the researcher by mail. Part of its attractiveness stems from its self-administered aspect: There are no interviewers to recruit, train, monitor, and compensate. Similarly, mailing lists are readily available from companies that specialize in this business, and it is possible to access very specific groups of target respondents. For example, it is possible to obtain a list of urban planners specializing in urban planning who operate settlement patterns and land management in cities larger than 500,000 people. Also, one may opt to purchase computer files, printed labels, or even labelled envelopes from these companies. In fact, some list companies will even provide insertion and mailing services. There are a number of companies that sell mailing lists, and most, if not all, have online purchase options. If you want to see an example, look at the guided tour on the USA Data Web site (www.usadata.com/).

On a per-mailed respondent basis, mail surveys are very inexpensive. In fact, they are almost always the least expensive survey method in this regard. But mail surveys incur all of the problems associated with not having an interviewer present. Despite the fact that the mail survey is described as “powerful, effective, and efficient” by the American Statistical

Association, the mail survey is plagued by two major problems. The first is **no response**, which refers to questionnaires that are not returned. The second is **self-selection bias**, which means that those who do respond are probably different from those who do not fill out the questionnaire and return it and, therefore, the sample gained through this method is not representative of the general population. To be sure, the mail survey is not the only survey method that suffers from non response and self-selection bias. Failures to respond are found in all types of surveys, and marketing researchers must be constantly alert to the possibilities that their final samples are somehow different from the original list of potential respondents because of some systematic tendency or latent pattern of response. Whatever the survey mode used, those who respond may be more involved with the product, they may have more education, they might be more or less dissatisfied, or they may even be more opinionated in general than the target population of concern. When informing clients of data collection alternatives, market researchers should inform them of the non response problems and biases inherent in each one being considered. For example, mail surveys are notorious for low response, and those respondents who do fill out and return a mail questionnaire are likely to be different from those who do not. At the same time, there are people who refuse to answer questions over the telephone, and consumers who like to shop are more likely to be encountered in mall-intercept interviews than are those who do not like to shop. Each data collection method has its own non response and bias considerations, and a conscientious researcher will help his or her client understand the dangers represented in the methods under consideration. Thus, non response and the subsequent danger of self-selection bias are greatest with mail surveys, for typically mail surveys of households achieve response rates of less than 20 percent.

5.3 Telephone survey

Information collected from the respondents by asking them questions on the phone is called as telephone interview or survey. Telephone surveying is defined as a systematic collection of data from a sample population using a standardized questionnaire. This method of collecting information consists in contacting respondents on telephone itself. It is not a very widely used method, but plays important part in industrial surveys, particularly in developed regions. The chief merits of such a system are:

1. It is more flexible in comparison to mailing method.
2. It is faster than other methods i.e., a quick way of obtaining information.
3. It is cheaper than personal interviewing method; here the cost per response is relatively low.
4. Recall is easy; call backs are simple and economical.
5. There is a higher rate of response than what we have in mailing method; the non-response is generally very low.
6. Replies can be recorded without causing embarrassment to respondents.
7. Interviewer can explain requirements more easily.
8. At times, access can be gained to respondents who otherwise cannot be contacted for one reason or the other.

9. No field staffs are required.
10. Representative and wider distribution of sample is possible.

But this system of collecting information is not free from demerits. Some of these may be highlighted as follows.

1. Little time is given to respondents for considered answers; interview period is not likely to exceed five minutes in most cases.
2. Surveys are restricted to respondents who have telephone facilities.
3. Extensive geographical coverage may get restricted by cost considerations.
4. It is not suitable for intensive surveys where comprehensive answers are required to various questions.
5. Possibility of the bias of the interviewer is relatively more.
6. Questions have to be short and to the point; probes are difficult to handle.

5.4 Home and Mall-Intercept survey

Just as the name implies, an **in-home interview** is conducted in the home of the respondent. Two important factors justify the use of in-home interviews. First, the researcher must believe that personal contact is essential to the success of the interview. Second, he or she must be convinced that the in-home environment is conducive to the questioning process.

With respect to the first factor, the survey may incorporate a set of advertisements the researcher wants viewed, it might require the respondent to see and touch a product, and the respondent may have to perform a complicated task such as sorting cards with brand names on them into piles, or it might be vital that the interviewer make visual confirmation of the respondent's qualifying characteristics or nonverbal cues. On the second factor, it is often believed that conducting an interview in the home greatly improves the quality of responses and facilitates the rapport between interviewer and interviewee. When a respondent is in a secure, comfortable environment, the likelihood of distraction is reduced, and it is believed that respondents take more care in responding to various questions. Also, most in-home interviews take considerable time, and by allowing the interview to take place in his or her home, the respondent is implicitly expecting it to take some time, certainly longer than would be expected for a telephone survey, for instance.

Although the in-home interview has important advantages, it has the significant disadvantage of cost. The expense of in-home interviewer travel is high, even for local surveys. Patterned after "man-on-the-street" interviews pioneered by opinion-polling companies and other "high-traffic" surveys conducted in settings where crowds of pedestrians pass by, the **mall-intercept interview** is one in which the respondent is encountered and questioned while he or she is visiting a shopping mall. A mall-intercept company generally has its offices located within a large shopping mall, usually one that draws from a regional rather than a local market area. Typically, the interview company negotiates exclusive rights to do interviews in

the mall and, thus, forces all marketing research companies that wish to do mall intercepts in that area to use that interview company's services. In any case, the travel costs are eliminated because the respondents incur the costs themselves by travelling to the mall.

Mall-intercept interviewing has acquired a major role as a survey method due to its ease of implementation. Shoppers are intercepted in the pedestrian traffic areas of shopping malls and either interviewed on the spot or asked to move to a permanent interviewing facility located in the mall office. Although some malls do not allow marketing research interviewing because they view it as a nuisance to shoppers, many do permit mall-intercept interviews and may rely on these data themselves to fine-tune their own marketing programs.

In addition to low cost, mall interviews have most of the advantages associated with in-home interviewing. Perhaps the most important advantage is the presence of an interviewer who can interact with the respondent. However, a few disadvantages are specifically associated with mall interviewing, and it is necessary to point them out here. First, sample representativeness is an issue, for most malls draw from a relatively small area in close proximity to their location. Some people shop at malls more frequently than others and, therefore, have a greater chance of being interviewed. Recent growth of non mall retailing concepts such as catalogs and stand-alone discounters such as Wal-Mart mean that more mall visitors are recreational shoppers rather than convenience-oriented shoppers, resulting in the need to scrutinize mall-intercept samples as to what consumer groups they actually represent. Also, many shoppers refuse to take part in mall interviews for various reasons. Nevertheless, special selection procedures called quotas are implemented.

A second shortcoming of mall-intercept interviewing is that a shopping mall does not have a comfortable home environment that is conducive to rapport and close attention to details. The respondents may feel uncomfortable because passersby stare at them; they may be pressed for time or otherwise preoccupied by various distractions outside the researcher's control. These factors may adversely affect the quality of the interview. As it is indicated earlier, some interview companies attempt to counter this problem by taking respondents to special interview rooms located in the interview company's mall offices. This procedure minimizes distractions and encourages respondents to be more relaxed.