first in terms of financial expenditure and later in terms of weight. The control questions, thus, introduce a cross-check to see whether the information collected is correct or not. Questions affecting the sentiments of respondents should be avoided. Adequate space for answers should be provided in the questionnaire to help editing and tabulation. There should always be provision for indications of uncertainty, e.g., "do not know," "no preference" and so on. Brief directions with regard to filling up the questionnaire should invariably be given in the questionnaire itself. Finally, the physical appearance of the questionnaire affects the cooperation the researcher receives from the recipients and as such an attractive looking questionnaire, particularly in mail surveys, is a plus point for enlisting cooperation. The quality of the paper, along with its colour, must be good so that it may attract the attention of recipients.

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, 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.

DIFFERENCE BETWEEN QUESTIONNAIRES AND SCHEDULES

Both questionnaire and schedule are popularly used methods of collecting data in research surveys. There is much resemblance in the nature of these two methods and this fact has made many people to remark that from a practical point of view, the two methods can be taken to be the same. But from the technical point of view there is difference between the two. The important points of difference are as under:

1. The questionnaire is generally sent through mail to informants to be answered as specified in a covering letter, but otherwise without further assistance from the sender. The schedule

is generally filled out by the research worker or the enumerator, who can interpret questions when necessary.

2. To collect data through questionnaire is relatively cheap and economical since we have to spend money only in preparing the questionnaire and in mailing the same to respondents. Here no field staff required. To collect data through schedules is relatively more expensive since considerable amount of money has to be spent in appointing enumerators and in importing training to them. Money is also spent in preparing schedules.
3. Non-response is usually high in case of questionnaire as many people do not respond and many return the questionnaire without answering all questions. Bias due to non-response often remains indeterminate. As against this, non-response is generally very low in case of schedules because these are filled by enumerators who are able to get answers to all questions. But there remains the danger of interviewer bias and cheating.
4. In case of questionnaire, it is not always clear as to who replies, but in case of schedule the identity of respondent is known.
5. The questionnaire method is likely to be very slow since many respondents do not return the questionnaire in time despite several reminders, but in case of schedules the information is collected well in time as they are filled in by enumerators.
6. Personal contact is generally not possible in case of the questionnaire method as questionnaires are sent to respondents by post who also in turn return the same by post. But in case of schedules direct personal contact is established with respondents.
7. Questionnaire method can be used only when respondents are literate and cooperative, but in case of schedules the information can be gathered even when the respondents happen to be illiterate.
8. Wider and more representative distribution of sample is possible under the questionnaire method, but in respect of schedules there usually remains the difficulty in sending enumerators over a relatively wider area.
9. Risk of collecting incomplete and wrong information is relatively more under the questionnaire method, particularly when people are unable to understand questions properly. But in case of schedules, the information collected is generally complete and accurate as enumerators can remove the difficulties, if any, faced by respondents in correctly understanding the questions. As a result, the information collected through schedules is relatively more accurate than that obtained through questionnaires.
10. The success of questionnaire method lies more on the quality of the questionnaire itself, but in the case of schedules much depends upon the honesty and competence of enumerators.
11. In order to attract the attention of respondents, the physical appearance of questionnaire must be quite attractive, but this may not be so in case of schedules as they are to be filled in by enumerators and not by respondents.
12. Along with schedules, observation method can also be used but such a thing is not possible while collecting data through questionnaires.

SOME OTHER METHODS OF DATA COLLECTION

Let us consider some other methods of data collection, particularly used by big business houses in modern times.

1. Warranty cards: Warranty cards are usually postal sized cards which are used by dealers of consumer durables to collect information regarding their products. The information sought is printed in the form of questions on the 'warranty cards' which is placed inside the package along with the product with a request to the consumer to fill in the card and post it back to the dealer.

2. Distributor or store audits: Distributor or store audits are performed by distributors as well as manufactures through their salesmen at regular intervals. Distributors get the retail stores audited through salesmen and use such information to estimate market size, market share, seasonal purchasing pattern and so on. The data are obtained in such audits not by questioning but by observation. For instance, in case of a grocery store audit, a sample of stores is visited periodically and data are recorded on inventories on hand either by observation or copying from store records. Store audits are invariably panel operations, for the derivation of sales estimates and compilation of sales trends by stores are their principal '*raison detre*'. The principal advantage of this method is that it offers the most efficient way of evaluating the effect on sales of variations of different techniques of in-store promotion.

3. Pantry audits: Pantry audit technique is used to estimate consumption of the basket of goods at the consumer level. In this type of audit, the investigator collects an inventory of types, quantities and prices of commodities consumed. Thus in pantry audit data are recorded from the examination of consumer's pantry. The usual objective in a pantry audit is to find out what types of consumers buy certain products and certain brands, the assumption being that the contents of the pantry accurately portray consumer's preferences. Quite often, pantry audits are supplemented by direct questioning relating to reasons and circumstances under which particular products were purchased in an attempt to relate these factors to purchasing habits. A pantry audit may or may not be set up as a panel operation, since a single visit is often considered sufficient to yield an accurate picture of consumers' preferences. An important limitation of pantry audit approach is that, at times, it may not be possible to identify consumers' preferences from the audit data alone, particularly when promotion devices produce a marked rise in sales.

4. Consumer panels: An extension of the pantry audit approach on a regular basis is known as 'consumer panel', where a set of consumers are arranged to come to an understanding to maintain detailed daily records of their consumption and the same is made available to investigator on demands. In other words, a consumer panel is essentially a sample of consumers who are interviewed repeatedly over a period of time. Mostly consume panels are of two types viz., the transitory consumer panel and the continuing consumer panel. *A transitory consumer panel* is set up to measure the effect of a particular phenomenon. Usually such a panel is conducted on a before-and-after-basis. Initial interviews are conducted before the phenomenon takes place to record the attitude of the consumer. A second set of interviews is carried out after the phenomenon has taken place to find out the consequent changes that might have occurred in the consumer's attitude. It is a favourite tool of advertising and of social research. A *continuing consumer panel* is often set up for an indefinite period with a view to collect data on a particular aspect of consumer behaviour over time, generally at periodic intervals or may be meant to serve as a general purpose panel for researchers on a variety of subjects. Such panels have been used in the area of consumer expenditure, public opinion and radio and TV listenership

among others. Most of these panels operate by mail. The representativeness of the panel relative to the population and the effect of panel membership on the information obtained after the two major problems associated with the use of this method of data collection.

5. Use of mechanical devices: The use of mechanical devices has been widely made to collect information by way of indirect means. Eye camera, Pupilometric camera, Psychogalvanometer, Motion picture camera and Audiometer are the principal devices so far developed and commonly used by modern big business houses, mostly in the developed world for the purpose of collecting the required information.

Eye cameras are designed to record the focus of eyes of a respondent on a specific portion of a sketch or diagram or written material. Such an information is useful in designing advertising material. Pupilometric cameras record dilation of the pupil as a result of a visual stimulus. The extent of dilation shows the degree of interest aroused by the stimulus. Psychogalvanometer is used for measuring the extent of body excitement as a result of the visual stimulus. Motion picture cameras can be used to record movement of body of a buyer while deciding to buy a consumer good from a shop or big store. Influence of packaging or the information given on the label would stimulate a buyer to perform certain physical movements which can easily be recorded by a hidden motion picture camera in the shop's four walls. Audiometers are used by some TV concerns to find out the type of programmes as well as stations preferred by people. A device is fitted in the television instrument itself to record these changes. Such data may be used to find out the market share of competing television stations.

6. Projective techniques: Projective techniques (or what are sometimes called as indirect interviewing techniques) for the collection of data have been developed by psychologists to use projections of respondents for inferring about underlying motives, urges, or intentions which are such that the respondent either resists to reveal them or is unable to figure out himself. In projective techniques the respondent in supplying information tends unconsciously to project his own attitudes or feelings on the subject under study. Projective techniques play an important role in motivational researches or in attitude surveys.

The use of these techniques requires intensive specialised training. In such techniques, the individual's responses to the stimulus-situation are not taken at their face value. The stimuli may arouse many different kinds of reactions. The nature of the stimuli and the way in which they are presented under these techniques do not clearly indicate the way in which the response is to be interpreted. The stimulus may be a photograph, a picture, an inkblot and so on. Responses to these stimuli are interpreted as indicating the individual's own view, his personality structure, his needs, tensions, etc. in the context of some pre-established psychological conceptualisation of what the individual's responses to the stimulus mean.

We may now briefly deal with the important projective techniques.

(i) *Word association tests:* These tests are used to extract information regarding such words which have maximum association. In this sort of test the respondent is asked to mention the first word that comes to mind, ostensibly without thinking, as the interviewer reads out each word from a list. If the interviewer says *cold*, the respondent may say *hot* and the like ones. The general technique is to use a list of as many as 50 to 100 words. Analysis of the matching words supplied by the respondents indicates whether the given word should be used for the contemplated purpose. The same idea is exploited in marketing research to find out the quality that is mostly associated to a brand of a product. A number of qualities of a product may be listed and informants may be asked to write

brand names possessing one or more of these. This technique is quick and easy to use, but yields reliable results when applied to words that are widely known and which possess essentially one type of meaning. This technique is frequently used in advertising research.

(ii) *Sentence completion tests:* These tests happen to be an extension of the technique of word association tests. Under this, informant may be asked to complete a sentence (such as: persons who wear Khadi are...) to find association of Khadi clothes with certain personality characteristics. Several sentences of this type might be put to the informant on the same subject. Analysis of replies from the same informant reveals his attitude toward that subject, and the combination of these attitudes of all the sample members is then taken to reflect the views of the population. This technique permits the testing not only of words (as in case of word association tests), but of ideas as well and thus, helps in developing hypotheses and in the construction of questionnaires. This technique is also quick and easy to use, but it often leads to analytical problems, particularly when the response happens to be multidimensional.

(iii) *Story completion tests:* Such tests are a step further wherein the researcher may contrive stories instead of sentences and ask the informants to complete them. The respondent is given just enough of story to focus his attention on a given subject and he is asked to supply a conclusion to the story.

(iv) *Verbal projection tests:* These are the tests wherein the respondent is asked to comment on or to explain what other people do. For example, why do people smoke? Answers may reveal the respondent's own motivations.

(v) *Pictorial techniques:* There are several pictorial techniques. The important ones are as follows:

(a) *Thematic apperception test (T.A.T.):* The TAT consists of a set of pictures (some of the pictures deal with the ordinary day-to-day events while others may be ambiguous pictures of unusual situations) that are shown to respondents who are asked to describe what they think the pictures represent. The replies of respondents constitute the basis for the investigator to draw inferences about their personality structure, attitudes, etc.

(b) *Rosenzweig test:* This test uses a cartoon format wherein we have a series of cartoons with words inserted in 'balloons' above. The respondent is asked to put his own words in an empty balloon space provided for the purpose in the picture. From what the respondents write in this fashion, the study of their attitudes can be made.

(c) *Rorschach test:* This test consists of ten cards having prints of inkblots. The design happens to be symmetrical but meaningless. The respondents are asked to describe what they perceive in such symmetrical inkblots and the responses are interpreted on the basis of some pre-determined psychological framework. This test is frequently used but the problem of validity still remains a major problem of this test.

(d) *Holtzman Inkblot Test (HIT):* This test from W.H. Holtzman is a modification of the Rorschach Test explained above. This test consists of 45 inkblot cards (and not 10 inkblots as we find in case of Rorschach Test) which are based on colour, movement, shading and other factors involved in inkblot perception. Only one response per card is obtained from the subject (or the respondent) and the responses of a subject are interpreted at three levels of form appropriateness. Form responses are interpreted for knowing the accuracy (F) or inaccuracy (F–) of respondent's percepts; shading and colour for ascertaining his affectional and emotional needs; and movement responses for assessing the dynamic aspects of his life.

Holtzman Inkblot Test or H.I.T. has several special features or advantages. For example, it elicits relatively constant number of responses per respondent. Secondly, it facilitates studying the responses of a respondent to different cards in the light of norms of each card instead of lumping them together. Thirdly, it elicits much more information from the respondent then is possible with merely 10 cards in Rorschach test; the 45 cards used in this test provide a variety of stimuli to the respondent and as such the range of responses elicited by the test is comparatively wider.

There are some limitations of this test as well. One difficulty that remains in using this test is that most of the respondents do not know the determinants of their perceptions, but for the researcher, who has to interpret the protocols of a subject and understand his personality (or attitude) through them, knowing the determinant of each of his response is a must. This fact emphasises that the test must be administered individually and a post-test inquiry must as well be conducted for knowing the nature and sources of responses and this limits the scope of HIT as a group test of personality. Not only this, "the usefulness of HIT for purposes of personal selection, vocational guidance, etc. is still to be established."[1]

In view of these limitations, some people have made certain changes in applying this test. For instance, Fisher and Cleveland in their approach for obtaining Barrier score of an individual's personality have developed a series of multiple choice items for 40 of HIT cards. Each of these cards is presented to the subject along with three acceptable choices [such as 'Knight in armour' (Barrier response), 'X-Ray' (Penetrating response) and 'Flower' (Neutral response)]. Subject taking the test is to check the choice he likes most, make a different mark against the one he likes least and leave the third choice blank. The number of barrier responses checked by him determines his barrier score on the test.

(e) *Tomkins-Horn picture arrangement test:* This test is designed for group administration. It consists of twenty-five plates, each containing three sketches that may be arranged in different ways to portray sequence of events. The respondent is asked to arrange them in a sequence which he considers as reasonable. The responses are interpreted as providing evidence confirming certain norms, respondent's attitudes, etc.

(vi) *Play techniques:* Under play techniques subjects are asked to improvise or act out a situation in which they have been assigned various roles. The researcher may observe such traits as hostility, dominance, sympathy, prejudice or the absence of such traits. These techniques have been used for knowing the attitudes of younger ones through manipulation of dolls. Dolls representing different racial groups are usually given to children who are allowed to play with them freely. The manner in which children organise dolls would indicate their attitude towards the class of persons represented by dolls. This is also known as *doll-play test*, and is used frequently in studies pertaining to sociology. The choice of colour, form, words, the sense of orderliness and other reactions may provide opportunities to infer deep-seated feelings.

(vii) *Quizzes, tests and examinations:* This is also a technique of extracting information regarding specific ability of candidates indirectly. In this procedure both long and short questions are framed to test through them the memorising and analytical ability of candidates.

(viii) *Sociometry:* Sociometry is a technique for describing the social relationships among individuals in a group. In an indirect way, sociometry attempts to describe attractions or repulsions between

[1] S.L. Dass, "*Personality Assessment Through Projective Movie Pictures*", p. 17.

individuals by asking them to indicate whom they would choose or reject in various situations. Thus, sociometry is a new technique of studying the underlying motives of respondents. "Under this an attempt is made to trace the flow of information amongst groups and then examine the ways in which new ideas are diffused. Sociograms are constructed to identify leaders and followers."[2] Sociograms are charts that depict the sociometric choices. There are many versions of the sociogram pattern and the reader is suggested to consult specialised references on sociometry for the purpose. This approach has been applied to the diffusion of ideas on drugs amongst medical practitioners.

7. 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.

8. Content-analysis: Content-analysis consists of analysing the contents of documentary materials such as books, magazines, newspapers and the contents of all other verbal materials which can be either spoken or printed. Content-analysis prior to 1940's was mostly quantitative analysis of documentary materials concerning certain characteristics that can be identified and counted. But since 1950's content-analysis is mostly qualitative analysis concerning the general import or message of the existing documents. "The difference is somewhat like that between a casual interview and depth interviewing."[3] Bernard Berelson's name is often associated with. the latter type of content-analysis. "Content-analysis is measurement through proportion…. Content analysis measures pervasiveness and that is sometimes an index of the intensity of the force."[4]

The analysis of content is a central activity whenever one is concerned with the study of the nature of the verbal materials. A review of research in any area, for instance, involves the analysis of the contents of research articles that have been published. The analysis may be at a relatively simple level or may be a subtle one. It is at a simple level when we pursue it on the basis of certain characteristics of the document or verbal materials that can be identified and counted (such as on the basis of major scientific concepts in a book). It is at a subtle level when researcher makes a study of the attitude, say of the press towards education by feature writers.

[2] G.B. Giles, *Marketing,* p. 40–41.

[3] Carter V. Good and Douglas E. Scates, *Methods of Research,* p. 670.

[4] *Ibid.,* p. 670.

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 are 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. In this connection Dr. A.L. Bowley very aptly observes that it is never safe to take published statistics at their face value without knowing their meaning and limitations and it is always necessary to criticise arguments that can be based on them.

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? (t) 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.

SELECTION OF APPROPRIATE METHOD FOR DATA COLLECTION

Thus, there are various methods of data collection. As such the researcher must judiciously select the method/methods 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. 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. Dr. A.L. Bowley's remark in this context is very appropriate when he says that "in collection of statistical data common sense is the chief requisite and experience the chief teacher."

CASE STUDY METHOD

Meaning: The case study method is a very popular form of qualitative analysis and involves a careful and complete observation of a social unit, be that unit a person, a family, an institution, a cultural group or even the entire community. It is a method of study in depth rather than breadth. The case study places more emphasis on the full analysis of a limited number of events or conditions and their interrelations. The case study deals with the processes that take place and their interrelationship. Thus, case study is essentially an intensive investigation of the particular unit under consideration. The object of the case study method is to locate the factors that account for the behaviour-patterns of the given unit as an integrated totality.

According to H. Odum, "The case study method is a technique by which individual factor whether it be an institution or just an episode in the life of an individual or a group is analysed in its relationship to any other in the group."[5] Thus, a fairly exhaustive study of a person (as to what he does and has done, what he thinks he does and had done and what he expects to do and says he ought to do) or group is called a life or case history. Burgess has used the words "the social microscope" for the case study method."[6] Pauline V. Young describes case study as "a comprehensive study of a social unit be that unit a person, a group, a social institution, a district or a community."[7] In brief, we can say that case study method is a form of qualitative analysis where in careful and complete observation of an individual or a situation or an institution is done; efforts are made to study each and every aspect of the concerning unit in minute details and then from case data generalisations and inferences are drawn.

Characteristics: The important characteristics of the case study method are as under:

1. Under this method the researcher can take one single social unit or more of such units for his study purpose; he may even take a situation to study the same comprehensively.
2. Here the selected unit is studied intensively i.e., it is studied in minute details. Generally, the study extends over a long period of time to ascertain the natural history of the unit so as to obtain enough information for drawing correct inferences.

[5] H. Odum, *An Introduction to Social Research*, p. 229.

[6] Burgess, *Research Methods in Sociology*, p. 26 in Georges Gurvitch and W.E. Moore (Eds.) *Twentieth Century Sociology*.

[7] Pauline V. Young, *Scientific Social Surveys and Research*, p. 247.

3. In the context of this method we make complete study of the social unit covering all facets. Through this method we try to understand the complex of factors that are operative within a social unit as an integrated totality.
4 Under this method the approach happens to be qualitative and not quantitative. Mere quantitative information is not collected. Every possible effort is made to collect information concerning all aspects of life. As such, case study deepens our perception and gives us a clear insight into life. For instance, under this method we not only study how many crimes a man has done but shall peep into the factors that forced him to commit crimes when we are making a case study of a man as a criminal. The objective of the study may be to suggest ways to reform the criminal.
5. In respect of the case study method an effort is made to know the mutual inter-relationship of causal factors.
6. Under case study method the behaviour pattern of the concerning unit is studied directly and not by an indirect and abstract approach.
7. Case study method results in fruitful hypotheses along with the data which may be helpful in testing them, and thus it enables the generalised knowledge to get richer and richer. In its absence, generalised social science may get handicapped.

Evolution and scope: The case study method is a widely used systematic field research technique in sociology these days. The credit for introducing this method to the field of social investigation goes to Frederic Le Play who used it as a hand-maiden to statistics in his studies of family budgets. Herbert Spencer was the first to use case material in his comparative study of different cultures. Dr. William Healy resorted to this method in his study of juvenile delinquency, and considered it as a better method over and above the mere use of statistical data. Similarly, anthropologists, historians, novelists and dramatists have used this method concerning problems pertaining to their areas of interests. Even management experts use case study methods for getting clues to several management problems. In brief, case study method is being used in several disciplines. Not only this, its use is increasing day by day.

Assumptions: The case study method is based on several assumptions. The important assumptions may be listed as follows:

(i) The assumption of uniformity in the basic human nature in spite of the fact that human behaviour may vary according to situations.
(ii) The assumption of studying the natural history of the unit concerned.
(iii) The assumption of comprehensive study of the unit concerned.

Major phases involved: Major phases involved in case study are as follows:

(i) Recognition and determination of the status of the phenomenon to be investigated or the unit of attention.
(ii) Collection of data, examination and history of the given phenomenon.
(iii) Diagnosis and identification of causal factors as a basis for remedial or developmental treatment.
(iv) Application of remedial measures i.e., treatment and therapy (this phase is often characterised as case work).

(v) Follow-up programme to determine effectiveness of the treatment applied.

Advantages: There are several advantages of the case study method that follow from the various characteristics outlined above. Mention may be made here of the important advantages.

(i) Being an exhaustive study of a social unit, the case study method enables us to understand fully the behaviour pattern of the concerned unit. In the words of Charles Horton Cooley, "case study deepens our perception and gives us a clearer insight into life…. It gets at behaviour directly and not by an indirect and abstract approach."

(ii) Through case study a researcher can obtain a real and enlightened record of personal experiences which would reveal man's inner strivings, tensions and motivations that drive him to action along with the forces that direct him to adopt a certain pattern of behaviour.

(iii) This method enables the researcher to trace out the natural history of the social unit and its relationship with the social factors and the forces involved in its surrounding environment.

(iv) It helps in formulating relevant hypotheses along with the data which may be helpful in testing them. Case studies, thus, enable the generalised knowledge to get richer and richer.

(v) The method facilitates intensive study of social units which is generally not possible if we use either the observation method or the method of collecting information through schedules. This is the reason why case study method is being frequently used, particularly in social researches.

(vi) Information collected under the case study method helps a lot to the researcher in the task of constructing the appropriate questionnaire or schedule for the said task requires thorough knowledge of the concerning universe.

(vii) The researcher can use one or more of the several research methods under the case study method depending upon the prevalent circumstances. In other words, the use of different methods such as depth interviews, questionnaires, documents, study reports of individuals, letters, and the like is possible under case study method.

(viii) Case study method has proved beneficial in determining the nature of units to be studied along with the nature of the universe. This is the reason why at times the case study method is alternatively known as "mode of organising data".

(ix) This method is a means to well understand the past of a social unit because of its emphasis of historical analysis. Besides, it is also a technique to suggest measures for improvement in the context of the present environment of the concerned social units.

(x) Case studies constitute the perfect type of sociological material as they represent a real record of personal experiences which very often escape the attention of most of the skilled researchers using other techniques.

(xi) Case study method enhances the experience of the researcher and this in turn increases his analysing ability and skill.

(xii) This method makes possible the study of social changes. On account of the minute study of the different facets of a social unit, the researcher can well understand the social change then and now. This also facilitates the drawing of inferences and helps in maintaining the continuity of the research process. In fact, it may be considered the gateway to and at the same time the final destination of abstract knowledge.

(xiii) Case study techniques are indispensable for therapeutic and administrative purposes. They are also of immense value in taking decisions regarding several management problems. Case data are quite useful for diagnosis, therapy and other practical case problems.

Limitations: Important limitations of the case study method may as well be highlighted.

(i) Case situations are seldom comparable and as such the information gathered in case studies is often not comparable. Since the subject under case study tells history in his own words, logical concepts and units of scientific classification have to be read into it or out of it by the investigator.

(ii) Read Bain does not consider the case data as significant scientific data since they do not provide knowledge of the "impersonal, universal, non-ethical, non-practical, repetitive aspects of phenomena."[8] Real information is often not collected because the subjectivity of the researcher does enter in the collection of information in a case study.

(iii) The danger of false generalisation is always there in view of the fact that no set rules are followed in collection of the information and only few units are studied.

(iv) It consumes more time and requires lot of expenditure. More time is needed under case study method since one studies the natural history cycles of social units and that too minutely.

(v) The case data are often vitiated because the subject, according to Read Bain, may write what he thinks the investigator wants; and the greater the rapport, the more subjective the whole process is.

(vi) Case study method is based on several assumptions which may not be very realistic at times, and as such the usefulness of case data is always subject to doubt.

(vii) Case study method can be used only in a limited sphere., it is not possible to use it in case of a big society. Sampling is also not possible under a case study method.

(viii) Response of the investigator is an important limitation of the case study method. He often thinks that he has full knowledge of the unit and can himself answer about it. In case the same is not true, then consequences follow. In fact, this is more the fault of the researcher rather than that of the case method.

Conclusion: Despite the above stated limitations, we find that case studies are being undertaken in several disciplines, particularly in sociology, as a tool of scientific research in view of the several advantages indicated earlier. Most of the limitations can be removed if researchers are always conscious of these and are well trained in the modern methods of collecting case data and in the scientific techniques of assembling, classifying and processing the same. Besides, case studies, in modern times, can be conducted in such a manner that the data are amenable to quantification and statistical treatment. Possibly, this is also the reason why case studies are becoming popular day by day.

Question

1. Enumerate the different methods of collecting data. Which one is the most suitable for conducting enquiry regarding family welfare programme in India? Explain its merits and demerits.

[8] Pauline V. Young, *Scientific social surveys and research,* p. 262.

2. "It is never safe to take published statistics at their face value without knowing their meaning and limitations." Elucidate this statement by enumerating and explaining the various points which you would consider before using any published data. Illustrate your answer by examples wherever possible.

3. Examine the merits and limitations of the observation method in collecting material. Illustrate your answer with suitable examples.

4. Describe some of the major projective techniques and evaluate their significance as tools of scientific social research.

5. How does the case study method differ from the survey method? Analyse the merits and limitations of case study method in sociological research.

6. Clearly explain the difference between collection of data through questionnaires and schedules.

7. Discuss interview as a technique of data collection.

8. Write short notes on:

(a) Depth interviews;

(b) Important aspects of a questionnaire;

(c) Pantry and store audits;

(d) Thematic Apperception Test;

(e) Holtzman Inkbolt Test.

9. What are the guiding considerations in the construction of questionnaire? Explain.

10. Critically examine the following:

(i) Interviews introduce more bias than does the use of questionnaire.

(ii) Data collection through projective techniques is considered relatively more reliable.

(iii) In collection of statistical data commonsense is the chief requisite and experience the chief teacher.

11. Distinguish between an experiment and survey. Explain fully the survey method of research.

[*M. Phi. (EAFM) Exam. 1987 Raj. Uni.*]

12. "Experimental method of research is not suitable in management field." Discuss, what are the problems in the introduction of this research design in business organisation?

[*M.B.A. (Part I) Exam. 1985 Raj. Uni.*]

Appendix (i)

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 his research project.
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 be 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.

Appendix (ii)

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 must be a man with self-restraint and self-discipline.
5. To the extent possible 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 to be observed by all so as to ensure reasonable uniformity in respect of all salient points in the study.

Appendix (iii)

Difference Between Survey and Experiment

The following points are noteworthy so far as difference between survey and experiment is concerned:

(i) Surveys are conducted in case of descriptive research studies where as experiments are a part of experimental research studies.

(ii) Survey-type research studies usually have larger samples because the percentage of responses generally happens to be low, as low as 20 to 30%, especially in mailed questionnaire studies. Thus, the survey method gathers data from a relatively large number of cases at a particular time; it is essentially cross-sectional. As against this, experimental studies generally need small samples.

(iii) Surveys are concerned with describing, recording, analysing and interpreting conditions that either exist or existed. The researcher does not manipulate the variable or arrange for events to happen. Surveys are only concerned with conditions or relationships that exist, opinions that are held, processes that are going on, effects that are evident or trends that are developing. They are primarily concerned with the present but at times do consider past events and influences as they relate to current conditions. Thus, in surveys, variables that exist or have already occurred are selected and observed.

Experimental research provides a systematic and logical method for answering the question, "What will happen if this is done when certain variables are carefully controlled or manipulated?" In fact, deliberate manipulation is a part of the experimental method. In an experiment, the researcher measures the effects of an experiment which he conducts intentionally.

(iv) Surveys are usually appropriate in case of social and behavioural sciences (because many types of behaviour that interest the researcher cannot be arranged in a realistic setting) where as experiments are mostly an essential feature of physical and natural sciences.

(v) Surveys are an example of field research where as experiments generally constitute an example of laboratory research.

(vi) Surveys are concerned with hypothesis formulation and testing the analysis of the relationship between non-manipulated variables. Experimentation provides a method of hypothesis testing. After experimenters define a problem, they propose a hypothesis. They then test the hypothesis and confirm or disconfirm it in the light of the controlled variable relationship that they have observed. The confirmation or rejection is always stated in terms of probability rather than certainty. Experimentation, thus, is the most sophisticated, exacting and powerful method for discovering and developing an organised body of knowledge. The ultimate purpose of experimentation is to generalise the variable relationships so that they may be applied outside the laboratory to a wider population of interest.*

(vii) Surveys may either be census or sample surveys. They may also be classified as social surveys, economic surveys or public opinion surveys. Whatever be their type, the method of data collection happens to be either observation, or interview or questionnaire/opinionnaire or some projective technique(s). Case study method can as well be used. But in case of experiments, data are collected from several readings of experiments.

(viii) In case of surveys, research design must be rigid, must make enough provision for protection against bias and must maximise reliability as the aim happens to be to obtain complete and accurate information. Research design in case of experimental studies, apart reducing bias and ensuring reliability, must permit drawing inferences about causality.

(ix) Possible relationships between the data and the unknowns in the universe can be studied through surveys where as experiments are meant to determine such relationships.

(x) Causal analysis is considered relatively more important in experiments where as in most social and business surveys our interest lies in understanding and controlling relationships between variables and as such correlation analysis is relatively more important in surveys.

* John W. Best and James V. Kahn, "*Research in Education*", 5th ed., Prentice-Hall of India Pvt. Ltd., New Delhi, 1986, p.111.

7

Processing and Analysis of Data

The data, after collection, has to be processed and analysed in accordance with the outline laid down for the purpose at the time of developing the research plan. This is essential for a scientific study and for ensuring that we have all relevant data for making contemplated comparisons and analysis. 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".[1] But there are persons (Selltiz, Jahoda and others) who do not like to make difference between processing and analysis. They opine that analysis of data in a general way involves a number of closely related operations which are performed with the purpose of summarising the collected data and organising these in such a manner that they answer the research question(s). We, however, shall prefer to observe the difference between the two terms as stated here in order to understand their implications more clearly.

PROCESSING OPERATIONS

With this brief introduction concerning the concepts of processing and analysis, we can now proceed with the explanation of all the processing operations.

1. Editing: Editing of data is a process of examining the collected raw data (specially in surveys) to detect errors and omissions and to correct these when possible. As a matter of fact, editing involves a careful scrutiny of the completed questionnaires and/or schedules. Editing is done to assure that the data are accurate, consistent with other facts gathered, uniformly entered, as completed as possible and have been well arranged to facilitate coding and tabulation.

With regard to points or stages at which editing should be done, one can talk of field editing and central editing. *Field editing* consists in the review of the reporting forms by the investigator for completing (translating or rewriting) what the latter has written in abbreviated and/or in illegible form

[1] G.B. Giles, *Marketing,* p. 44.

at the time of recording the respondents' responses. This type of editing is necessary in view of the fact that individual writing styles often can be difficult for others to decipher. This sort of editing should be done as soon as possible after the interview, preferably on the very day or on the next day. While doing field editing, the investigator must restrain himself and must not correct errors of omission by simply guessing what the informant would have said if the question had been asked.

Central editing should take place when all forms or schedules have been completed and returned to the office. This type of editing implies that all forms should get a thorough editing by a single editor in a small study and by a team of editors in case of a large inquiry. Editor(s) may correct the obvious errors such as an entry in the wrong place, entry recorded in months when it should have been recorded in weeks, and the like. In case of inappropriate on missing replies, the editor can sometimes determine the proper answer by reviewing the other information in the schedule. At times, the respondent can be contacted for clarification. The editor must strike out the answer if the same is inappropriate and he has no basis for determining the correct answer or the response. In such a case an editing entry of 'no answer' is called for. All the wrong replies, which are quite obvious, must be dropped from the final results, especially in the context of mail surveys.

Editors must keep in view several points while performing their work: (a) They should be familiar with instructions given to the interviewers and coders as well as with the editing instructions supplied to them for the purpose. (b) While crossing out an original entry for one reason or another, they should just draw a single line on it so that the same may remain legible. (c) They must make entries (if any) on the form in some distinctive colur and that too in a standardised form. (d) They should initial all answers which they change or supply. (e) Editor's initials and the date of editing should be placed on each completed form or schedule.

2. Coding: Coding refers to the process of assigning numerals or other symbols to answers so that responses can be put into a limited number of categories or classes. Such classes should be appropriate to the research problem under consideration. They must also possess the characteristic of exhaustiveness (i.e., there must be a class for every data item) and also that of mutual exclusively which means that a specific answer can be placed in one and only one cell in a given category set. Another rule to be observed is that of unidimensionality by which is meant that every class is defined in terms of only one concept.

Coding is necessary for efficient analysis and through it the several replies may be reduced to a small number of classes which contain the critical information required for analysis. Coding decisions should usually be taken at the designing stage of the questionnaire. This makes it possible to precode the questionnaire choices and which in turn is helpful for computer tabulation as one can straight forward key punch from the original questionnaires. But in case of hand coding some standard method may be used. One such standard method is to code in the margin with a coloured pencil. The other method can be to transcribe the data from the questionnaire to a coding sheet. Whatever method is adopted, one should see that coding errors are altogether eliminated or reduced to the minimum level.

3. Classification: Most research studies result in a large volume of raw data which must be reduced into homogeneous groups if we are to get meaningful relationships. This fact necessitates classification of data which happens to be the process of arranging data in groups or classes on the basis of common characteristics. Data having a common characteristic are placed in one class and in this

way the entire data get divided into a number of groups or classes. Classification can be one of the following two types, depending upon the nature of the phenomenon involved:

(a) *Classification according to attributes:* As stated above, data are classified on the basis of common characteristics which can either be descriptive (such as literacy, sex, honesty, etc.) or numerical (such as weight, height, income, etc.). Descriptive characteristics refer to qualitative phenomenon which cannot be measured quantitatively; only their presence or absence in an individual item can be noticed. Data obtained this way on the basis of certain attributes are known as *statistics of attributes* and their classification is said to be classification according to attributes.

Such classification can be simple classification or manifold classification. In simple classification we consider only one attribute and divide the universe into two classes—one class consisting of items possessing the given attribute and the other class consisting of items which do not possess the given attribute. But in manifold classification we consider two or more attributes simultaneously, and divide that data into a number of classes (total number of classes of final order is given by 2^n, where n = number of attributes considered).* Whenever data are classified according to attributes, the researcher must see that the attributes are defined in such a manner that there is least possibility of any doubt/ambiguity concerning the said attributes.

(b) *Classification according to class-intervals:* Unlike descriptive characteristics, the numerical characteristics refer to quantitative phenomenon which can be measured through some statistical units. Data relating to income, production, age, weight, etc. come under this category. Such data are known as *statistics of variables* and are classified on the basis of class intervals. For instance, persons whose incomes, say, are within Rs 201 to Rs 400 can form one group, those whose incomes are within Rs 401 to Rs 600 can form another group and so on. In this way the entire data may be divided into a number of groups or classes or what are usually called, 'class-intervals.' Each group of class-interval, thus, has an upper limit as well as a lower limit which are known as class limits. The difference between the two class limits is known as class magnitude. We may have classes with equal class magnitudes or with unequal class magnitudes. The number of items which fall in a given class is known as the frequency of the given class. All the classes or groups, with their respective frequencies taken together and put in the form of a table, are described as group frequency distribution or simply frequency distribution. Classification according to class intervals usually involves the following three main problems:

(i) How may classes should be there? What should be their magnitudes?

There can be no specific answer with regard to the number of classes. The decision about this calls for skill and experience of the researcher. However, the objective should be to display the data in such a way as to make it meaningful for the analyst. Typically, we may have 5 to 15 classes. With regard to the second part of the question, we can say that, to the extent possible, class-intervals should be of equal magnitudes, but in some cases unequal magnitudes may result in better classification. Hence the

* Classes of the final order are those classes developed on the basis of '*n*' attributes considered. For example, if attributes *A* and *B* are studied and their presence is denoted by *A* and *B* respectively and absence by *a* and *b* respectively, then we have four classes of final order viz., class *AB*, class *Ab*, class *aB*, and class *ab*.

researcher's objective judgement plays an important part in this connection. Multiples of 2, 5 and 10 are generally preferred while determining class magnitudes. Some statisticians adopt the following formula, suggested by H.A. Sturges, determining the size of class interval:

$$i = R/(1 + 3.3 \log N)$$

where

i = size of class interval;

R = Range (i.e., difference between the values of the largest item and smallest item among the given items);

N = Number of items to be grouped.

It should also be kept in mind that in case one or two or very few items have very high or very low values, one may use what are known as open-ended intervals in the overall frequency distribution. Such intervals may be expressed like under Rs 500 or Rs 10001 and over. Such intervals are generally not desirable, but often cannot be avoided. The researcher must always remain conscious of this fact while deciding the issue of the total number of class intervals in which the data are to be classified.

(ii) How to choose class limits?

While choosing class limits, the researcher must take into consideration the criterion that the mid-point (generally worked out first by taking the sum of the upper limit and lower limit of a class and then divide this sum by 2) of a class-interval and the actual average of items of that class interval should remain as close to each other as possible. Consistent with this, the class limits should be located at multiples of 2, 5, 10, 20, 100 and such other figures. Class limits may generally be stated in any of the following forms:

Exclusive type class intervals: They are usually stated as follows:

10–20

20–30

30–40

40–50

The above intervals should be read as under:

10 and under 20

20 and under 30

30 and under 40

40 and under 50

Thus, under the exclusive type class intervals, the items whose values are equal to the upper limit of a class are grouped in the next higher class. For example, an item whose value is exactly 30 would be put in 30–40 class interval and not in 20–30 class interval. In simple words, we can say that under exclusive type class intervals, the upper limit of a class interval is excluded and items with values less than the upper limit (but not less than the lower limit) are put in the given class interval.

Inclusive type class intervals: They are usually stated as follows:

11–20
21–30
31–40
41–50

In inclusive type class intervals the upper limit of a class interval is also included in the concerning class interval. Thus, an item whose value is 20 will be put in 11–20 class interval. The stated upper limit of the class interval 11–20 is 20 but the real limit is 20.99999 and as such 11–20 class interval really means 11 and under 21.

When the phenomenon under consideration happens to be a discrete one (i.e., can be measured and stated only in integers), then we should adopt inclusive type classification. But when the phenomenon happens to be a continuous one capable of being measured in fractions as well, we can use exclusive type class intervals.*

(iii) How to determine the frequency of each class?

This can be done either by tally sheets or by mechanical aids. Under the technique of tally sheet, the class-groups are written on a sheet of paper (commonly known as the tally sheet) and for each item a stroke (usually a small vertical line) is marked against the class group in which it falls. The general practice is that after every four small vertical lines in a class group, the fifth line for the item falling in the same group, is indicated as horizontal line through the said four lines and the resulting flower (IIII) represents five items. All this facilitates the counting of items in each one of the class groups. An illustrative tally sheet can be shown as under:

Table 7.1: An Illustrative Tally Sheet for Determining the Number of 70 Families in Different Income Groups

Income groups (Rupees)	*Tally mark*	*Number of families or (Class frequency)*
Below 400	𝍸 𝍸 III	13
401–800	𝍸 𝍸 𝍸 𝍸	20
801–1200	𝍸 𝍸 II	12
1201–1600	𝍸 𝍸 𝍸 III	18
1601 and above	𝍸 II	7
Total		70

Alternatively, class frequencies can be determined, specially in case of large inquires and surveys, by mechanical aids i.e., with the help of machines viz., sorting machines that are available for the purpose. Some machines are hand operated, whereas other work with electricity. There are machines

* The stated limits of class intervals are different than true limits. We should use true or real limits keeping in view the nature of the given phenomenon.

which can sort out cards at a speed of something like 25000 cards per hour. This method is fast but expensive.

4. Tabulation: When a mass of data has been assembled, it becomes necessary for the researcher to arrange the same in some kind of concise and logical order. This procedure is referred to as tabulation. Thus, tabulation is the process of summarising raw data and displaying the same in compact form (i.e., in the form of statistical tables) for further analysis. In a broader sense, tabulation is an orderly arrangement of data in columns and rows.

Tabulation is essential because of the following reasons.

1. It conserves space and reduces explanatory and descriptive statement to a minimum.
2. It facilitates the process of comparison.
3. It facilitates the summation of items and the detection of errors and omissions.
4. It provides a basis for various statistical computations.

Tabulation can be done by hand or by mechanical or electronic devices. The choice depends on the size and type of study, cost considerations, time pressures and the availaibility of tabulating machines or computers. In relatively large inquiries, we may use mechanical or computer tabulation if other factors are favourable and necessary facilities are available. Hand tabulation is usually preferred in case of small inquiries where the number of questionnaires is small and they are of relatively short length. Hand tabulation may be done using the direct tally, the list and tally or the card sort and count methods. When there are simple codes, it is feasible to tally directly from the questionnaire. Under this method, the codes are written on a sheet of paper, called tally sheet, and for each response a stroke is marked against the code in which it falls. Usually after every four strokes against a particular code, the fifth response is indicated by drawing a diagonal or horizontal line through the strokes. These groups of five are easy to count and the data are sorted against each code conveniently. In the listing method, the code responses may be transcribed onto a large work-sheet, allowing a line for each questionnaire. This way a large number of questionnaires can be listed on one work sheet. Tallies are then made for each question. The card sorting method is the most flexible hand tabulation. In this method the data are recorded on special cards of convenient size and shape with a series of holes. Each hole stands for a code and when cards are stacked, a needle passes through particular hole representing a particular code. These cards are then separated and counted. In this way frequencies of various codes can be found out by the repetition of this technique. We can as well use the mechanical devices or the computer facility for tabulation purpose in case we want quick results, our budget permits their use and we have a large volume of straight forward tabulation involving a number of cross-breaks.

Tabulation may also be classified as simple and complex tabulation. The former type of tabulation gives information about one or more groups of independent questions, whereas the latter type of tabulation shows the division of data in two or more categories and as such is deigned to give information concerning one or more sets of inter-related questions. Simple tabulation generally results in one-way tables which supply answers to questions about one characteristic of data only. As against this, complex tabulation usually results in two-way tables (which give information about two inter-related characteristics of data), three-way tables (giving information about three interrelated characteristics of data) or still higher order tables, also known as manifold tables, which supply

information about several interrelated characteristics of data. Two-way tables, three-way tables or manifold tables are all examples of what is sometimes described as cross tabulation.

Generally accepted principles of tabulation: Such principles of tabulation, particularly of constructing statistical tables, can be briefly states as follows:*

1. Every table should have a clear, concise and adequate title so as to make the table intelligible without reference to the text and this title should always be placed just above the body of the table.
2. Every table should be given a distinct number to facilitate easy reference.
3. The column headings (captions) and the row headings (stubs) of the table should be clear and brief.
4. The units of measurement under each heading or sub-heading must always be indicated.
5. Explanatory footnotes, if any, concerning the table should be placed directly beneath the table, along with the reference symbols used in the table.
6. Source or sources from where the data in the table have been obtained must be indicated just below the table.
7. Usually the columns are separated from one another by lines which make the table more readable and attractive. Lines are always drawn at the top and bottom of the table and below the captions.
8. There should be thick lines to separate the data under one class from the data under another class and the lines separating the sub-divisions of the classes should be comparatively thin lines.
9. The columns may be numbered to facilitate reference.
10. Those columns whose data are to be compared should be kept side by side. Similarly, percentages and/or averages must also be kept close to the data.
11. It is generally considered better to approximate figures before tabulation as the same would reduce unnecessary details in the table itself.
12. In order to emphasise the relative significance of certain categories, different kinds of type, spacing and indentations may be used.
13. It is important that all column figures be properly aligned. Decimal points and (+) or (–) signs should be in perfect alignment.
14. Abbreviations should be avoided to the extent possible and ditto marks should not be used in the table.
15. Miscellaneous and exceptional items, if any, should be usually placed in the last row of the table.
16. Table should be made as logical, clear, accurate and simple as possible. If the data happen to be very large, they should not be crowded in a single table for that would make the table unwieldy and inconvenient.
17. Total of rows should normally be placed in the extreme right column and that of columns should be placed at the bottom.

* All these points constitute the characteristics of a good table.

18. The arrangement of the categories in a table may be chronological, geographical, alphabetical or according to magnitude to facilitate comparison. Above all, the table must suit the needs and requirements of an investigation.

SOME PROBLEMS IN PROCESSING

We can take up the following two problems of processing the data for analytical purposes:

(a) *The problem concerning "Don't know" (or DK) responses:* While processing the data, the researcher often comes across some responses that are difficult to handle. One category of such responses may be 'Don't Know Response' or simply DK response. When the DK response group is small, it is of little significance. But when it is relatively big, it becomes a matter of major concern in which case the question arises: Is the question which elicited DK response useless? The answer depends on two points viz., the respondent actually may not know the answer or the researcher may fail in obtaining the appropriate information. In the first case the concerned question is said to be alright and DK response is taken as legitimate DK response. But in the second case, DK response is more likely to be a failure of the questioning process.

How DK responses are to be dealt with by researchers? The best way is to design better type of questions. Good rapport of interviewers with respondents will result in minimising DK responses. But what about the DK responses that have already taken place? One way to tackle this issue is to estimate the allocation of DK answers from other data in the questionnaire. The other way is to keep DK responses as a separate category in tabulation where we can consider it as a separate reply category if DK responses happen to be legitimate, otherwise we should let the reader make his own decision. Yet another way is to assume that DK responses occur more or less randomly and as such we may distribute them among the other answers in the ratio in which the latter have occurred. Similar results will be achieved if all DK replies are excluded from tabulation and that too without inflating the actual number of other responses.

(b) *Use or percentages:* Percentages are often used in data presentation for they simplify numbers, reducing all of them to a 0 to 100 range. Through the use of percentages, the data are reduced in the standard form with base equal to 100 which fact facilitates relative comparisons. While using percentages, the following rules should be kept in view by researchers:

1. Two or more percentages must not be averaged unless each is weighted by the group size from which it has been derived.
2. Use of too large percentages should be avoided, since a large percentage is difficult to understand and tends to confuse, defeating the very purpose for which percentages are used.
3. Percentages hide the base from which they have been computed. If this is not kept in view, the real differences may not be correctly read.
4. Percentage decreases can never exceed 100 per cent and as such for calculating the percentage of decrease, the higher figure should invariably be taken as the base.
5. Percentages should generally be worked out in the direction of the causal-factor in case of two-dimension tables and for this purpose we must select the more significant factor out of the two given factors as the causal factor.

ELEMENTS/TYPES OF ANALYSIS

As stated earlier, by analysis we mean the computation of certain indices or measures along with searching for patterns of relationship that exist among the data groups. Analysis, particularly in case of survey or experimental data, involves estimating the values of unknown parameters of the population and testing of hypotheses for drawing inferences. Analysis may, therefore, be categorised as descriptive analysis and inferential analysis (Inferential analysis is often known as statistical analysis). "*Descriptive analysis* is largely the study of distributions of one variable. This study provides us with profiles of companies, work groups, persons and other subjects on any of a multiple of characteristics such as size. Composition, efficiency, preferences, etc."[2]. this sort of analysis may be in respect of one variable (described as unidimensional analysis), or in respect of two variables (described as bivariate analysis) or in respect of more than two variables (described as multivariate analysis). In this context we work out various measures that show the size and shape of a distribution(s) along with the study of measuring relationships between two or more variables.

We may as well talk of correlation analysis and causal analysis. *Correlation analysis* studies the joint variation of two or more variables for determining the amount of correlation between two or more variables. *Causal analysis* is concerned with the study of how one or more variables affect changes in another variable. It is thus a study of functional relationships existing between two or more variables. This analysis can be termed as regression analysis. Causal analysis is considered relatively more important in experimental researches, whereas in most social and business researches our interest lies in understanding and controlling relationships between variables then with determining causes *per se* and as such we consider correlation analysis as relatively more important.

In modern times, with the availability of computer facilities, there has been a rapid development of *multivariate analysis* which may be defined as "all statistical methods which simultaneously analyse more than two variables on a sample of observations"[3]. Usually the following analyses[*] are involved when we make a reference of multivariate analysis:

(a) *Multiple regression analysis:* This analysis is adopted when the researcher has one dependent variable which is presumed to be a function of two or more independent variables. The objective of this analysis is to make a prediction about the dependent variable based on its covariance with all the concerned independent variables.

(b) *Multiple discriminant analysis:* This analysis is appropriate when the researcher has a single dependent variable that cannot be measured, but can be classified into two or more groups on the basis of some attribute. The object of this analysis happens to be to predict an entity's possibility of belonging to a particular group based on several predictor variables.

(c) *Multivariate analysis of variance* (*or multi-ANOVA*): This analysis is an extension of two-way ANOVA, wherein the ratio of among group variance to within group variance is worked out on a set of variables.

(d) *Canonical analysis:* This analysis can be used in case of both measurable and non-measurable variables for the purpose of simultaneously predicting a set of dependent variables from their joint covariance with a set of independent variables.

[2] C. William Emory, *Business Research Methods*, p. 356.

[3] Jagdish N. Sheth, "The Multivariate Revolution in Marketing Research", *Journal of Marketing*, Vol. 35, No. 1 (Jan. 1971), pp. 13–19.

[*] Readers are referred to standard texts for more details about these analyses.

Inferential analysis is concerned with the various tests of significance for testing hypotheses in order to determine with what validity data can be said to indicate some conclusion or conclusions. It is also concerned with the estimation of population values. It is mainly on the basis of inferential analysis that the task of interpretation (i.e., the task of drawing inferences and conclusions) is performed.

STATISTICS IN RESEARCH

The role of statistics in research is to function as a tool in designing research, analysing its data and drawing conclusions therefrom. Most research studies result in a large volume of raw data which must be suitably reduced so that the same can be read easily and can be used for further analysis. Clearly the science of statistics cannot be ignored by any research worker, even though he may not have occasion to use statistical methods in all their details and ramifications. Classification and tabulation, as stated earlier, achieve this objective to some extent, but we have to go a step further and develop certain indices or measures to summarise the collected/classified data. Only after this we can adopt the process of generalisation from small groups (i.e., samples) to population. If fact, there are two major areas of statistics viz., descriptive statistics and inferential statistics. *Descriptive statistics* concern the development of certain indices from the raw data, whereas inferential statistics concern with the process of generalisation. *Inferential statistics* are also known as sampling statistics and are mainly concerned with two major type of problems: (i) the estimation of population parameters, and (ii) the testing of statistical hypotheses.

The important statistical measures[*] that are used to summarise the survey/research data are:

(1) measures of central tendency or statistical averages; (2) measures of dispersion; (3) measures of asymmetry (skewness); (4) measures of relationship; and (5) other measures.

Amongst the measures of central tendency, the three most important ones are the arithmetic average or mean, median and mode. Geometric mean and harmonic mean are also sometimes used.

From among the measures of dispersion, variance, and its square root—the standard deviation are the most often used measures. Other measures such as mean deviation, range, etc. are also used. For comparison purpose, we use mostly the coefficient of standard deviation or the coefficient of variation.

In respect of the measures of skewness and kurtosis, we mostly use the first measure of skewness based on mean and mode or on mean and median. Other measures of skewness, based on quartiles or on the methods of moments, are also used sometimes. Kurtosis is also used to measure the peakedness of the curve of the frequency distribution.

Amongst the measures of relationship, Karl Pearson's coefficient of correlation is the frequently used measure in case of statistics of variables, whereas Yule's coefficient of association is used in case of statistics of attributes. Multiple correlation coefficient, partial correlation coefficient, regression analysis, etc., are other important measures often used by a researcher.

Index numbers, analysis of time series, coefficient of contingency, etc., are other measures that may as well be used by a researcher, depending upon the nature of the problem under study.

We give below a brief outline of some important measures (our of the above listed measures) often used in the context of research studies.

[*] One may read any standard text book on statistical methods for details about these measures.

MEASURES OF CENTRAL TENDENCY

Measures of central tendency (or statistical averages) tell us the point about which items have a tendency to cluster. Such a measure is considered as the most representative figure for the entire mass of data. Measure of central tendency is also known as statistical average. Mean, median and mode are the most popular averages. *Mean*, also known as arithmetic average, is the most common measure of central tendency and may be defined as the value which we get by dividing the total of the values of various given items in a series by the total number of items. we can work it out as under:

$$\text{Mean (or } \bar{X})^{*} = \frac{\sum X_i}{n} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

where $\bar{X}$ = The symbol we use for mean (pronounced as X bar)

$\sum$ = Symbol for summation

X_i = Value of the ith item X, $i = 1, 2, \ldots, n$

n = total number of items

In case of a frequency distribution, we can work out mean in this way:

$$\bar{X} = \frac{\sum f_i X_i}{\sum f_i} = \frac{f_1 X_1 + f_2 X_2 + \ldots + f_n X_n}{f_1 + f_2 + \ldots + f_n = n}$$

Sometimes, instead of calculating the simple mean, as stated above, we may workout the weighted mean for a realistic average. The weighted mean can be worked out as follows:

$$\bar{X}_w = \frac{\sum w_i X_i}{\sum w_i}$$

where $\bar{X}_w$ = Weighted item

w_i = weight of ith item X

X_i = value of the ith item X

Mean is the simplest measurement of central tendency and is a widely used measure. Its chief use consists in summarising the essential features of a series and in enabling data to be compared. It is amenable to algebraic treatment and is used in further statistical calculations. It is a relatively stable measure of central tendency. But it suffers from some limitations viz., it is unduly affected by extreme items; it may not coincide with the actual value of an item in a series, and it may lead to wrong impressions, particularly when the item values are not given with the average. However, mean is better than other averages, specially in economic and social studies where direct quantitative measurements are possible.

Median is the value of the middle item of series when it is arranged in ascending or descending order of magnitude. It divides the series into two halves; in one half all items are less than median, whereas in the other half all items have values higher than median. If the values of the items arranged in the ascending order are: 60, 74, 80, 90, 95, 100, then the value of the 4th item viz., 88 is the value of median. We can also write thus:

* If we use assumed average A, then mean would be worked out as under:

$\bar{X} = A + \frac{\sum(X_i - A)}{n}$ or $\bar{X} = A + \frac{\sum f_i(X_i - A)}{\sum f_i}$, in case of frequency distribution. This is also known as short cut method of finding $\bar{X}$.

$$\text{Median } (M) = \text{Value of } \left(\frac{n+1}{2}\right)\text{th item}$$

Median is a positional average and is used only in the context of qualitative phenomena, for example, in estimating intelligence, etc., which are often encountered in sociological fields. Median is not useful where items need to be assigned relative importance and weights. It is not frequently used in sampling statistics.

Mode is the most commonly or frequently occurring value in a series. The mode in a distribution is that item around which there is maximum concentration. In general, mode is the size of the item which has the maximum frequency, but at items such an item may not be mode on account of the effect of the frequencies of the neighbouring items. Like median, mode is a positional average and is not affected by the values of extreme items. it is, therefore, useful in all situations where we want to eliminate the effect of extreme variations. Mode is particularly useful in the study of popular sizes. For example, a manufacturer of shoes is usually interested in finding out the size most in demand so that he may manufacture a larger quantity of that size. In other words, he wants a modal size to be determined for median or mean size would not serve his purpose. but there are certain limitations of mode as well. For example, it is not amenable to algebraic treatment and sometimes remains indeterminate when we have two or more model values in a series. It is considered unsuitable in cases where we want to give relative importance to items under consideration.

Geometric mean is also useful under certain conditions. It is defined as the *n*th root of the product of the values of *n* times in a given series. Symbolically, we can put it thus:

$$\text{Geometric mean (or G.M.)} = \sqrt[n]{\pi X_i}$$

$$= \sqrt[n]{X_1 \cdot X_2 \cdot X_3 \ldots X_n}$$

where

G.M. = geometric mean,

n = number of items.

X_i = ith value of the variable X

π = conventional product notation

For instance, the geometric mean of the numbers, 4, 6, and 9 is worked out as

$$\text{G.M.} = \sqrt[3]{4.6.9}$$

$$= 6$$

The most frequently used application of this average is in the determination of average per cent of change i.e., it is often used in the preparation of index numbers or when we deal in ratios.

Harmonic mean is defined as the reciprocal of the average of reciprocals of the values of items of a series. Symbolically, we can express it as under:

$$\text{Harmonic mean (H.M.)} = \text{Rec.}\frac{\sum \text{Rec} X_i}{n}$$

$$= \text{Rec.}\frac{\text{Rec.}\, X_1 + \text{Rec.}\, X_2 + \ldots + \text{Rec.}\, X_n}{n}$$

where

H.M. = Harmonic mean

Rec. = Reciprocal

X_i = ith value of the variable X

n = number of items

For instance, the harmonic mean of the numbers 4, 5, and 10 is worked out as

$$\text{H.M.} = \text{Rec}\frac{1/4 + 1/5 + 1/10}{3} = \text{Rec}\ \frac{\frac{15 + 12 + 6}{60}}{3}$$

$$= \text{Rec}\left(\frac{33}{60} \times \frac{1}{3}\right) = \frac{60}{11} = 5.45$$

Harmonic mean is of limited application, particularly in cases where time and rate are involved. The harmonic mean gives largest weight to the smallest item and smallest weight to the largest item. As such it is used in cases like time and motion study where time is variable and distance constant.

From what has been stated above, we can say that there are several types of statistical averages. Researcher has to make a choice for some average. There are no hard and fast rules for the selection of a particular average in statistical analysis for the selection of an average mostly depends on the nature, type of objectives of the research study. One particular type of average cannot be taken as appropriate for all types of studies. The chief characteristics and the limitations of the various averages must be kept in view; discriminate use of average is very essential for sound statistical analysis.

MEASURES OF DISPERSION

An averages can represent a series only as best as a single figure can, but it certainly cannot reveal the entire story of any phenomenon under study. Specially it fails to give any idea about the scatter of the values of items of a variable in the series around the true value of average. In order to measure this scatter, statistical devices called measures of dispersion are calculated. Important measures of dispersion are (a) range, (b) mean deviation, and (c) standard deviation.

(a) *Range* is the simplest possible measure of dispersion and is defined as the difference between the values of the extreme items of a series. Thus,

$$\text{Range} = \begin{pmatrix}\text{Highest value of an}\\ \text{item in a series}\end{pmatrix} - \begin{pmatrix}\text{Lowest value of an}\\ \text{item in a series}\end{pmatrix}$$

The utility of range is that it gives an idea of the variability very quickly, but the drawback is that range is affected very greatly by fluctuations of sampling. Its value is never stable, being based on only two values of the variable. As such, range is mostly used as a rough measure of variability and is not considered as an appropriate measure in serious research studies.

(b) *Mean deviation* is the average of difference of the values of items from some average of the series. Such a difference is technically described as deviation. In calculating mean deviation we ignore the minus sign of deviations while taking their total for obtaining the mean deviation. Mean deviation is, thus, obtained as under: