UNIT III

Subject Name	Subject Code	Semester	Prepared by
Research Methods for Business	**18BBA55S**	**V**	**Dr Prabhu R Assistant Professor**

COLLECTION AND PROCESSING OF DATA

Data constitute the foundation of any scientific analysis and interpretation. There is a subtle difference between data and information. While data are just facts or recorded measures of certain phenomena, information refer to a body of facts in a format that facilities decision making or in a context that defines relationship between pieces of data. Data can be obtained from two important sources *viz.* (i) secondary source and (ii) primary source. Depending on the source, the data will be secondary or primary.

Secondary data: When an investigator (the student researcher) uses the data which have already been collected by others, such data are called secondary data.

Classification of secondary data

i.) Common external sources

Secondary data obtained from books, journals, reports, govt. publications, publications of organizations, professional bodies, project reports, media, commercial sources (market share data, stock market sources etc) and etc. are known as common external sources.

ii.) Internal and proprietary data: These data are generally available in the routine business in various departments like finance, production, personnel, sales, R.D. department and etc. Data on employees' salary from payroll, sales amounts from sales journal, raw materials from stock registers, labor and manufacturing expenses from production records and cash receipts from cash books are internal data. Data seen in balance sheet and profit –loss account are the major internal data available for projects in finance.

iii.) Tertiary sources:

These are interpretations of a secondary resource but generally represented by indexes, bibliographic and other finding aids such as internet search engines.

Saunders *et al* (2006) classified the secondary data into three categories viz., Documentary, Multiple source and Survey

Secondary data

Documentary		**Multiple source**		**Survey**		
Written Materials	**Non-Written Materials**	**Area based**	**Time-series based**	**Censuses**	**Continuous and regular surveys**	**Ad hoc surveys**
Examples	**Examples**	**Examples**	**Examples**	**Examples**	**Examples**	**Examples**
Organisations' databases, such as personnel or production. Organisations' communications	Media accounts including TV and radio Voice recordings	Financial Times country reports Government publications	Industry statistics and reports Government publications	Governments' censuses Census of population	Government: Family spending Labour market trends Organisation:	Government surveys Organisations' surveys
such as emails, letters, memos Organisations' Websites Reports and minutes of committees Journals Newspapers Diaries Interview transcripts	Video recordings	Books Journals	Publications by international organizations Books Journals	Census of Employment	BMRB International's Target group Index Employee attitude surveys	Academics' surveys

Table: Types of Secondary data (Saunders et al, 2006)

Features of secondary sources of data:

- Readymade and readily available
- Consist of data over which the researcher has no control
- Do not require the direct involvement of the researcher

Before using the secondary data, the investigator must examine the following aspects:

- Whether the data are suitable for the purpose of investigation
- Whether the data are adequate for the investigation
- Whether the data are reliable.

Objectives of collecting secondary data

The main objectives of collecting secondary data are:
Fact finding : To collect descriptive information to support decision making.
Model building : To specify relationships between variables based on secondary data, sometimes using descriptive or predictive equations.
Data mining : To dig through computers and analyse volumes of data to discover patterns about customers products and activities.

Advantages of secondary source of data

The advantages of utilizing secondary data are many.

Cost and time : Good quality data can be accessed in a short period for a tiny fraction of the resources involved in carrying out data collection by the researcher himself / herself
Opportunity for longitudinal analysis: Secondary data can offer the opportunity for longitudinal research .
Subgroup or subset analysis : When large samples are the source of secondary data, there is opportunity to study a sizeable sub-groups of individuals or subsets of questions.
Opportunity for cross – cultural analysis: Secondary sources of comparable data from two

or more countries with cultural differences provides possible models for conducting cross-cultural research.

More time for data analysis: As data collection is time consuming, there may not be adequate time for analysis and interpretation. When secondary courses of data are used [which can be collected in a short period] there is plenty of time for analysis and interpretation.

Reanalysis may offer new interpretations: Analysis of data which has already been analyzed may yield further insights into the problem. New theories, which were not available to earlier researchers, may help arrive at different interpretation. An alternative method of quantitative data analysis not employed by the original research may give different interpretation to the secondary source of data.

Limitations of secondary data

Inspite of many advantages, usage of secondary data is beset with some limitations also *Lack of familiarity with data*: With data collected by others it is not easy to get familiarized with the range of variables, the ways in which variables were coded and various aspects of the organization of the data. The period of familiarization with the data of secondary sources can be quite substantial.

Complexity of the data: Some of the secondary data sources are huge with large no. of respondents as well as variables. Data would have been collected at different levels [individual, group or organizational]

Quality of data: The researcher does not have control over the quality of data. Though many data sources of government agencies and reputed research organizations are quite dependable, the quality of data should not be taken for granted.

Appropriateness: The data from secondary sources may not meet the requirements of the researcher.

Absence of key variables: Since the data of secondary sources were collected for the purpose of a project of someone else there is no guarantee that all the key variables needed for a new researcher will be available in the data.

Hence, before using the secondary data the researcher must seek favourable answers to the following questions:

- How pertinent are the secondary data to the present investigation?
- Does the subject matter match with the problem definition?
- Are the data appropriate to apply to the population of interest?
- Is the time period consistent with present need?
- Do the units of measurement appear in correct format?
- Do the data adequately cover the details of the present study?

The most common problems associated with secondary data are:

- outdated information
- variation in definition of terms.
- different units of measurement and

Suitability of secondary data

Saunders *et al* (2011) included an exhaustive checklist covering the overall suitability and costs and benefits of secondary data:

Overall Suitability

- Does the data set contain the information required to answer the research questions(s) and meet the objectives?
- Do the measures used match those required?
- Is the data set a proxy for the data that is really needed?
- Does the data set cover the population that is the subject of the research?
- Does the data set cover the geographical area that is the subject of your research?
- Can data about the population that is the subject of the research be separated from unwanted data?
- Are the data for the right time period or sufficiently upto date?
- Are data available for all variables that are required to answer the research question(s) and meet the objectives?

Primary Data:

Primary data are collected directly by the researcher from the original sources. In primary data collection the researcher can collect the required data according to his research needs. He can collect them when he wants them and in the form he needs them. Of course, primary data collection is costly and time-consuming. But in several cases secondary data may not be available necessitating collection of primary data.

Methods of primary data collection:

The important methods are observation, experimentation, simulation and survey (interviewing, self-administered questionnaire and mail survey)

Observation: Scientific observation may be defined as a systematic viewing of a specific phenomenon in its proper setting for the purpose of gathering data for a particular study. Observation is not only seeing and hearing, it includes perceiving as well.

Observation becomes scientific when it serves a formulated research purpose.

- It is planned deliberately
- It is recorded systematically and
- It is subjected to checks and controls on validity and reliability
- It should answer the questions who, what, when, how and where.

Characteristics: Observation as a method of data collection, has certain characteristics:

- It is both a physical and a mental activity
- It is selective
- It is purposive, not casual
- It captures the natural context
- It grasps the significant events and occurrences

Types of observation: Observations are classified in different ways:

- Depending on the investigator's role it may be participant observation or non-participant observation
- In terms of the mode of observation it could be direct observation or indirect observation
- Considering the rigour of the system adopted it could be controlled observation or uncontrolled observation
- Based on the type of analysis, observation could be non behavioural observation and behavioural observation.

There are different types of observation

i. Participant observation: Here, the observer is a part of the phenomenon or group which is observed and he acts as both an observer and a participant to gain first hand knowledge. For

ii. Non-participant observation: In this observation the observer stands aloof and records his observation without actually participating in the phenomenon. This method calls for skill in recording observations without getting noticed.

iii. Direct observation: It is a straight forward attempt to observe and record what naturally occurs. The investigator observes the phenomenon personally when it takes place without creating an artificial situation the investigator observes the phenomenon personally when it takes place.

iv. Indirect observation(or mechanical observation): Here, the physical presence of the observer is not required and the recording is done by mechanical, photographic or electronicdevices.

v. Controlled Observation: In this method the observational techniques are standardized with maximum control over extrinsic and intrinsic variables by adopting experimental design and recording the observations systematically. In a typical controlled observation the decisions on what, how and when to observe will be explicit and clear. It is primarily used for inferring causality and testing causal hypothesis.

vi. Uncontrolled observation: This method does not involve control over extrinsic and intrinsic variables. It is primarily used for descriptive research.

vii. Non-behavioural observation: This includes record analysis, physical condition analysis and process or activity analysis.

Record analysis: This involves analysis of historical or current records and public or private records. They may be written, printed, sound-recorded, photographed or videotaped.

viii. Behavioural observation: This includes nonverbal analysis, linguistic analysis, extra linguistic analysis and spatial analysis.

Nonverbal behaviour is related to body movement, motor expressions, eye blink and glances. Body movements indicate a worker's interest or boredom, anger or pleasure. Motor expressions such as facial movements show the emotional state of a person; eye blink rates indicate interests in visuals and exchanged glances are indications of interpersonal behaviour.

Extra linguistic behaviour is an important means of communication related to vocal [pitch, loudness and temper], temporal [rate of speaking, duration of utterance and rhythm], interaction [tendencies to interrupt, dominate or inhibit] and verbal stylistic [vocabulary, pronunciation dialect and expressions.

Spacial relationship involves how a person relates physically to others. Known as *proxemics*, spacial relationship is concerned with how people maintain discrete distances between themselves and others while speaking and interacting.

Advantages of observation methods:

It is possible to study behaviour as it occurs. There is no need to ask the respondents about their opinion, attitude and etc.

- The data describe the observed phenomena as they occur in their natural settings
- It is the only method of studying subjects who are unable to articulate meaningfully (children, tribals, low level workers and etc)
- Contextual background of behaviour can be analysed. *i.e.* what one says and behaves can be compared.
- The investigator does not trouble the respondents to collect information
- For continuous observations over longer periods mechanical devices can be used

Limitations of observation methods:

- Past events or activities cannot be studied
- It is not suitable to study opinions and attitudes of the respondents
- The investigator has to wait for the event to occur
- Direct presence of the investigator for observation may not be possible in all cases
- Observation method is a slow and expensive process
- There is the risk of imposing an inappropriate framework on the setting being observed
- It is difficult to get an overall picture as observation generates lots of bits of data
- If neglects the context within which behaviour takes place.

Experimentation: It is a research process used to study the causal relationship between variables. It aims at studying the effect of an independent variable on a dependent variable keeping other extraneous variables under control. Experimentation is described in the Chapter on Research Designs.

Simulation: It is a process of conducting experiments on a symbolic model representing a phenomenon.

Types of simulation differ from situation to situation.

Man stimulation: It is a game played by people in a laboratory setting to simulate people in real life.

Computer simulation: This is an operational model programmed to generate a sequence of interactions. This type of simulation requires precise definitions, storing of large amounts of data on the system to be analysed and programming of the analysis.

Man-Computer Simulation: In this type, persons play the role of decision-makers while the computer is responsive to the players' activities. In this game, the computer is provided with a script.

*Organizational simulation:*This involves creating an artificial situation in which individual or group behavior can be observed. It is almost similar to a laboratory experiment, except that it does not seek to control participants activities as much as in an experimental research design. In this simulation, participants are given greater freedom to act according to their judgment and to make decisions and their actions are observed.

Simulation can be employed in behavioural problems, social problems, political problems, economic problems, war strategies, business problems and etc.

In the case of business problems such as determination of proper order quantities, production scheduling, designing complex distribution system, waiting line problems in transport services, maintenance scheduling in factories, airline and bus service, assembly line scheduling, consumer behaviour prediction, financial forecasting, introduction of new products, control system and etc simulation can be employed.

Surveys

Survey is a research technique in which primary data/ information are collected from a sample of people through a questionnaire or interview. It is a method of data collection based on communication with respondents. Respondent is a person who answers an interviewer's questions or provides answers to written questions in an interview schedule or self-administered questionnaire survey.

The popular methods employed in survey are interview [face-to-face, or telephone interview] and using a schedule, questionnaire and mailed questionnaire.

Interviewing: Interview is a common occurrence in social life such as job interviews, media interviews, social work interviews, police interviews, appraisal interviews and of course research interviews. These different kinds of interviews have some common features in eliciting information from the interviewee, formality and explicitness to the conduct of interview. In business research interview, the aim of the interview is to elicit information from the interviewee. In survey research, the interviewee is frequently called respondent.

Interview is defined as a two-way systematic conversation between the researcher and the respondent, initiated for obtaining information relevant to a specific study.

- Interview is a prominent method of data collection
- It involves not only conversation but also learning from the respondents' gestures, facial expressions and etc.
- It requires fact-to-face contact or contact over phone (now a days video conferencing)
- It demands special skills
- It may be structured or an unstructured
- It may be used either as a main method or as a supplementary one in studies of persons

- It is the only suitable method for gathering information from illiterate or less educated respondents
- It is useful for collecting a wide range of data from factual demographic data to highly personal and intimate information relating to the respondents opinions, attitudes, values, beliefs, past experience and future intentions.
- It is inevitable where probing is necessary
- It is superior to other data gathering models as people are usually more willing to talk than to write.
- It adds flesh to statistical information
- It permits the investigator to seek clarifications

Advantages of interviewing:

- With a well conceived schedule personal interview can obtain a great deal of information by prolonging the interview.
- Personal interview improves the percentage of response and the quality of information received.
- The interviewer can gather supplement information like the living conditions and environment of the respondents as participation is high
- The interviewer can use special scoring device to improve the quality of interviewing
- Props and visual aids can be used to elicit appropriate answers.
- The accuracy and dependability of the responses of the respondents can be checked by observation and probing
- Interview is flexible and adaptable to individual situations.
- There is opportunity for feedback from respondents
- There is scope for probing. That is, if the response is brief or unclear probing through verbal prompts, encourages respondents to answer adequately.

Limitations of interviewing:

- It is costly in terms of both money and time

- It may be adversely affected by the interviewer's mode of questioning / interactions and incorrect reading
- Respondents' faulty perception / memory and inability to articulate may influence the interview.
- In the face-to-face interview, the respondents may be reluctant to divulge information on personal matters and financial aspects
- Recording information may be a distraction to both he respondent and interviewer affecting the thread of conversation
- Interviewing is a highly skilled job and training of interviewers is a long and costly process.

Types of interview

i. Structured or directive interview:

- It is made with a detailed standardized schedule
- Same questions are put to all the respondents in the same order
- Each question is asked in the same way in each interview to promote measurement reliability
- It is used for large scale formalized surveys

Advantages:

- Data from one interview to the next one are easily comparable
- Greater precision is achieved
- Attention is not diverted to extraneous, irrelevant or time consuming conversation

Limitations

- It tends to lose the spontaneity of natural conversation
- Investigators' own biases may override the respondents views
- The scope of exploration is limited

ii. Semi-structured interview: It typically refers to a context in which the interviewer has a series of questions that are in the general form of an interview schedule but he is able to vary the sequence of questions, The questions are frequently more general in the frame of reference than found in a structured interview schedule. Further, the interviewer has some latitude to ask additional questions.

iii. Unstructured or Non-directive interview

- The interviewee is encouraged to talk freely with a minimum of prompting or guiding
- Instead of a detailed pre-planned interview only a broad interview guide is used
- The interviewer avoids channelising the interview direction
- It is more useful in case studies rather than in surveys.
- It is particularly useful in exploratory research where the lines of investigation are not clearly defined

Limitations

- Data obtained may not be comparable and hence not suitable for surveys
- Time may be wasted in unproductive conversation
- Coding and classifying of responses require more time
- It calls for greater skill than the formal survey interview

vi. Focused interview:

- It is a semi-structured interview where the investigator attempts to focus the discussion on the actual effects of a given experience to which the respondents have been exposed

- The focus is on the subjective experiences of the respondents such as their attitudes, emotional responses and etc. to particular situations.
- The situation is analysed prior to the interview and an interview guide related o the research hypothesis is used.

v. Clinical interview:

It is similar to the focused interview but with a subtle difference. While the focused interview is concerned with the effects of a specific objective, clinical interview is concerned with broad underlying feelings, motivations or life experiences.

vi. Depth interview:

It is relatively unstructured, extensive interview employed in the primary stages of a research process.

- It is an intensive and searching interview aiming at studying the respondents' opinion, emotions or convictions.
- It deliberately aims to elicit unconscious as well as extremely personal feelings and emotions
- It is generally a lengthy procedure designed to encourage free expression of one's feelings
- It requires real probing questions
- The interviewer should totally avoid advising or showing disagreement. Instead he/she should use encouraging expressions to motivate the respondent to continue discussion
- This interview requires much more training in interpersonal skills than structured interview

vii. Group Interview: It is a method of collecting data using a single interviewer with more than one participant [respondent] – Based on the size, a group is classified as dyad [two people] triad [three people] mini group [2-6 people] or super group [up to 20 people].

viii. Focus group interview: Focus group is a panel of people [typically 6 -10 participants] led by a trained moderator as interviewer who meet for 1 to 2 hours. In focus group interview, the moderator uses group dynamics to exchange ideas, feelings and experiences on a specific topic. Focus group interviews involve free association, picture/ photo sort, role play writing and drawing in addition to projective techniques.

ix. Telephonic interview:

It is contacting respondent by telephone to collect data/information in a survey.

It is useful when

- the population is composed of persons whose names are listed in the telephone directory
- the study requires responses to 5 or 6 simple questions
- the study is to be completed in a short period
- the study is interesting or important to the respondents
- the respondents are wide spread
- the respondents cannot afford to spend more time in answering a questionnaire

Advantages

- Telephone interviewing is less expensive
- Time taken is short
- As there is no fact-to-face contact, interviewer's bias is reduced resulting in quality responses
- Telephone call backs are easier than personal interview callbacks.
- It is less demanding on the interviewe

Disadvantages

- As population is restricted to persons in the telephone directory the study may be incomplete
- The responses cannot be elaborative
- Respondents' body language cannot be assessed
- Establishing rapport will not be easy
- As the identity of the interviewer is not confirmed, there may be suspicion resulting in poor response.
- Visual or complex questions cannot be used
- Respondents can terminate the interview even if the interviewer wants to continue
- Involvement of respondents is poor

x. Projective techniques: Projective technique is an indirect means of questioning that enables a respondent to 'project' beliefs and feelings onto a third person, an object or a situation. Since researchers are often looking for hidden or suppressed meanings, projective techniques can be used within the interview structures.

Self-administered questionnaire

It is a questionnaire filled in by the respondent rather than by the interviewer. If the questionnaire is designed to be filled to by the interviewer, it becomes an interview schedule.

Mail survey (postal / e-mail)

This method involves mailing the self-administered questionnaires to the respondents to get their responses. Other formats of mail survey are fax survey (using fax machines), e-mail survey and internet survey(posting the questionnaire on a website).

Advantages:

- Mail survey is less costly than personal interview
- It can cover extensive geographical areas

- It is useful in contacting busy executives who generally avoid interviews
- The respondents can fill up the questionnaire leisurely
- Due to anonymity responses may be free from inhibition
- As there is no personal contact, interviewer's bias is absent

Disadvantages

- It is limited only to literate population
- Response rate is generally low and hence the sample may not be representative
- The causes for inadequate or non-responses cannot be explained
- If the questionnaire is incomplete it cannot be rectified.

Steps to improve response rates to mailed questionnaire:

Because the response to mailed questionnaire is not encouraging in many cases, certain steps are suggested.

- Writing a good covering letter explaining the reasons for the research, why it isimportant, and why the respondent has been selected. There should also be a mention about sponsorship, if any. It must be invariably mentioned that confidentially of the information received will be maintained and that the information will not be used for purposes other than the present research.
- Postal questionnaires should always be accompanied by a stamped addressed envelope or return postage.
- If the response is not received within a reasonable period, 2 or 3 gentle reminders are to be sent at reasonable internals along with copies of questionnaire.
- Preferably the questionnaire must be shorter
- There must be clear instructions to fill up the questionnaire and the layout must be simple but alternative

Tools for data collection

The various methods of data collection involve the use of appropriate tools or instruments. These tools / instruments facilitate data collection

a.) Schedule: It is a proforma containing a set of questions which are asked and filled in by an interviewer personally [face-to-face]

i. Observation schedule or observationnaire

- This is a form used to record the observations
- The items to be observed are determined in advance and grouped into appropriate categories
- The items are structured with possible alternatives
- Space is provided against each unit observation

ii. Interview guide

- It is used for non-directive and depth interviews
- It contains only the broad topics or areas to be covered in the interview
- It serves as a suggestive reference or prompter during interview
- It aids in focusing attention on salient points relating to the study
- It helps in securing comparable data in different interviews by the same or different interviewers.

iii. Interview schedule and mailed questionnaire

- Both these tools are widely used in surveys
- Both contain lists of questions to be answered by the respondents
- While schedule is filled out by the interviewer, questionnaire is completed by the respondent

b.) Opinionaire

It is similar to interview schedule / mailed questionnaire. It is used for studying opinions of people. Wording of statements may differ.

c.) Inventory

It is essentially a list that the respondent is asked to mark or check in a particular way regarding personal qualities.

Construction of self-completion or mail/postal questionnaire or schedule:

A questionnaire / schedule should have provision to collect personal information of the respondents apart from the information on the study under reference. The personal information generally pertain to (i) Gender (ii) Age group (iii) Marital status (iv) Educational background (v) Occupation (vi) Salary / Income (vii) Experience and etc., depending on the nature of study. This information is essential to differentiate the opinion based on gender, age, income and etc of the respondents.

When surveys are conducted among businessmen / business houses, the data on the type of business, turnover, number of employees, dealers' network and etc. are collected.

The questions included may be open-ended, closed-end questions, dichotomous, multiple-choice questions or declarative.

i. Open-ended questions: These are unstructured ones providing free scope to the respondents to reply with their own choice of words and ideas. *(e.g.)* What is your opinion on the present trend in HR? What are your comments on the Govt. initiatives in improving FDI?

Open-ended questions are helpful in initiating discussion or for a in-depth investigation. The major limitation is that it is difficult for compilation, classification and analysis of the responses.

Advantages and disadvantages of open-ended questions

Advantages

- Respondents can answer in their own terms and not forced to answer in a particular way.
- Unusual response, not contemplated by the researcher may be derived.
- Respondents' levels of knowledge and understanding the research questions can be tapped. The importance of the research question for the respondents can also be explored.
- Open questions are useful to explore new areas in which the researcher has limited knowledge.

Disadvantages

- Time consuming as the respondents are likely to talk longer if the research topic interests them
- Coding of answers is time-consuming
- Post-coding can be an unreliable process, because it can introduce the variability in coding of answers resulting in measurement error (and hence lack of validity)
- As answers to open-questions need more efforts on the part of respondents, many prospective respondents may put off answering the questionnaire.
- There may be variability between interviewers in the recording of answers.

ii. Closed-end questions: These are structured ones with two or more alternative responses for the respondent to choose. They generally contain standardized answers. They are simple to administer and easy to compile and analyse. As the alternatives are designed with reference to the requirements of the study, the chances of securing relevant answers are better.

Advantages of closed question:

- It is easy to process the answers as pre-codes are placed to the side of the fixed-choice answers.
- Closed questions improve the comparability of answers making it easier to show the relationship between variables and to make comparisons between respondents
- Availability of answers along with the closed questions may help clarify the situation to the respondents.

- As the respondents are not expected to write extensively and instead have to place ticks or circle answers closed questions are easier and quicker to complete.
- Only relevant answers are

obtained The limitations of open-ended

questions

- The response is forced in terms of the investigator rather than that of the respondent. That is, there is loss of spontaneity in the respondents' answers.
- The respondent is led to choose a response even when he has no knowledge of it
- Closed questions may be irritating when the respondents are not able to find a choice that they feel appropriate
- The limited alternatives may not cover all the view points
- Different respondents may interpret the same words and statements differently.

iii. Dichotomous or two-choice questions: A dichotomous question can be answered either as 'yes' or 'no'. Here, there is no choice for the respondent to say 'undecided', 'do not know', 'sometimes' or 'partially'.

iv. Multiple choice questions: These questions contain more than 2 alternatives

(e.g.) Why do you prefer a particular brand of mobile hand-set?

- Price is reasonable
- Handy
- Sleek
- More features
- Others (please specify)

As the alternatives may not be all-inclusive, 'others (please specify)' is included to provide for any other option.

v. Declarative questions: It is a type of multiple choice questions in which a series of statements are given and the respondent is asked to select the one which represents his view.

Designing a self-completion questionnaire

The questionnaire must be designed in such a way that respondents answer the questions willingly and with ease.

Avoiding cramping of presentation: Though it is advisable to make the questionnaire as short as possible, it must not look cramped. Reducing margins and the space between questions and putting questions too close make the questionnaire cramped.

Presenting clearly: The questionnaire must have a layout that is easy on the eye and facilitates the answering of all questions. A variety of print styles (different fonts, print sizes, bold, italics and capitals) can enhance the appearance but must be used in a consistent way. It must be ensured that there is one style for general instructions, one for specific Instructions, one for closed-ended questions etc. Mixing print styles can be confusing.

Arranging fixed answers: As most of the questions in a self-completion questionnaire are likely to be of the closed type, the question is whether to arrange the fixed answers vertically or horizontally. Many writers prefer a vertical format whenever possible because confusion can arise when a horizontal type is used. For example, the respondent is asked to answer a particular question and the answers are given horizontally as very good-good –fair-poor-very poor. When the questionnaire is filled up in a hurry, the required tick will be placed in the wrong space such as indicating 'good' when 'fair' is the intended answer.

The desirable format is vertical as shown below:

Very good	-	5
Good	-	4
Fair	-	3
Poor	-	2
Very poor	-	1

In the vertical format not only is there less ambiguity about where tick is placed, the coding is easier.

Horizontal format as shown below is not desirable.

Very Good-Good-Fair-Poor-Very Poor - 5 4 3 2 1.

However, when there is a battery of questions with identical answer formats as in a Likert scale vertical format will take of too much space.

Giving clear instructions: The instructions must be clear whether the respondents are to place a tick or circle or underline the appropriate answer or are they supposed to delete the inappropriate answers. In certain cases the respondents may choose more than one answer. If it is not acceptable it should be indicated "Please choose the one answer that best represents your views by placing a tick in the appropriate box". If it is not made clear and if some of the respondents select more than one answer, then the question is to be treated as unanswered. If it is acceptable to have more than one choice, then the instruction should be, "Please choose all answers that represent your views by placing ticks in the appropriate boxes".

Keeping questions and answers together: It is a simple and obvious requirement that a question should not be split and given in two separate pages. Further, the answers must be given by the side of the questions.

Questionnaire vs Schedule

The main differences between questionnaire and schedule in collecting data/ information are given in table 8.3

Questionnaire	Schedule
➢ Filled in by the respondent	➢ Filled up by researcher/field workers
➢ Response is low as many may not return the questionnaire	➢ Response is high as data are collected in person
➢ In mailed questionnaire identity of respondent is not known	➢ Researcher/field workers know who is responding
➢ Literacy level is important	➢ Literacy level is not a major concern
➢ Wider distribution is possible	➢ Difficult to contact many
➢ May be the data are less accurate and incomplete	➢ Data are likely to be more accurate and complete

➢ Success depends on the type of questions	➢ Success depends on the person who collects data
➢ Physical appearance of questionnaire is important	➢ Not so important
➢ Cost is less	➢ Cost is more due to hiring field workers

Table: Questionnaire vs Schedule

PROCESSING OF DATA

Prior to any sort of statistical analysis for interpretation, the data collected from experiments, observations and surveys and etc are processed. Data processing includes identification of variables, hypothetical relationship, if any, among the variables and tentative research hypothesis.

Steps in data processing:

The important steps in data processing are:

- Identifying the types of information
- Editing the data
- Coding of data in the case of large volume of data and where secrecy is to be maintained
- Classifying and tabulation of data

i. Identifying the types of information:

The information collected is identified whether they are quantitative and qualitative in nature. All the qualitative information is converted into quantitative data. For instance, the opinion collected in respect of overall job satisfaction of employees on a 5-point Likert type scale is quantified appropriately (Strongly Agree gets 5 points, Agree 4 points, No idea 3 points, Disagree 2 points and Strongly Disagree 1 point).

ii. Editing:

Editing is the process of detecting and correcting of errors and omissions. Editing is done to make the data ready for coding and transfer to data storage. Editing is done at field level in large projects, before the data go for central editing. A thorough editing is done by central editing.

Editing is done the identify the errors and correct them. Editing is mainly done to check for completeness, accuracy and uniformity.

a. Completeness: The information gathered are scrutinized whether all the questions in a questionnaire / opinionaire / inventory are answered. In the case of observation it is checked whether all the relevant observations are made by the researcher as per the predetermined structure. In the case of minor omissions the researcher / editor can fill it up based on the available data. If the missing information is of vital importance to the study, the respondents are contracted again. In experimental designs, 'missing plot technique' is employed to find out the missing data.

b. Accuracy: A random check may be helpful to verify the accuracy of data. Along with accuracy, consistency in responses can also be ascertained. Irresponsible responses are deleted. One way of checking the accuracy/consistency is to repeat the question in different format, either as a negative or positive statement.

c. Uniformity: To elicit meaningful responses and uniform interpretation the questions formulated should be unambiguous. In the case of opinionaire or inventory the statements must be either only positive or only negative. If positive and negative statements are mixed up 'reverse scoring' is resorted to during analysis. In the positive statement the opinion Strongly Agree, Agree, No idea, Disagree or Strongly Disagree get points 5, 4, 3, 2 and 1 respectively. In the case of negative statement the points will be 1,2,3,4 and 5 accordingly.

Some of the problems encountered are inappropriate respondents, incomplete interview, lack of understanding, inconsistency, illegibility and fictitious interview.

iii. Coding of Data:

Coding involves assigning numbers or other symbols to answers so that the responses can be grouped note a limited no of categories. In comparatively small research projects, such as MBA summer project, where the number of respondents in a survey is around 200 – 250 or the number of companies or the number of financial years taken is few, coding is not necessary. But when the number of variables and respondents are many coding is done. Further, coding is essential when

the researcher feels that some sort of secrecy is required when the data are subjected to analysis. The coding could be numeric or alphabetic.

(e.g.) Gender: Male many be codified as M or 1 and Female as F or 2

Occupation: Salaried as S or 1

Business as B or 2

Professional as P or 3

Retired as R or 4

Others as O or 5

Categorization of data must be

- appropriate to the research problem / purpose
- exhaustive
- mutually exclusive and
- derived from one classification dimension

iv. Classification and Tabulation:

Classification must be appropriate depending on the available data in relation to the objectives of the study. Suppose the researcher desires to know the satisfaction level of the different age groups of employees, then the employees are classified as young (<30 years) middle aged (31 to 45 years) and old (> 45 years). In the survey, where the age difference is not perceptible as in the case of BPO organization, classification based on age becomes irrelevant.

Classification should be exhaustive i.e., all the categories are taken into consideration. When the researcher wants to classify the respondents based on martial status generally they are classified as married, or unmarried. Here, the widower, widow or separated cannot find a place. To overcome this category 'others' is included.

The categories must be mutually exhaustive. For example, while classifying the respondents based on income it could be < R.10,000, Rs.10001 to 20000, Rs. 20001 to 30000 and

more than Rs. 30000. It should not be < Rs.10000, Rs. 10000 – 20000 and 20000 to 30000, which is not mutually exclusive.

Tabulation

Tabulation is the process of summarizing raw data into meaningful categories and presenting them for a quick understanding. Tabulation is a pre-requisite for statistical analysis.

Components of a table:

- Heading - Table number, title of the table and designation of units
- Body – Stub-head – headings of all rows or blocks of sub-items
- Body head – headings of all columns or main captions and sub-captions
- Field /Body – cells in rows and columns
- Notations - foot notes and sources wherever applicable.

Types of Table: Based on the number of dimensions or variables tables are classified as frequency table, cross tabulation and contingency tables.

Frequency table: It is a simple tabulation that indicate frequency with which respondents give a particular answers or respondents are classified.

Example: Respondents' response to a particular question. Have you ever visited Brookefields in Coimbatore?

Suppose 500 persons have been contacted and 376 have visited and 124 have not visited. The frequency table is presented as in table 8.5

S.No	Response	Frequency (no. of people)
1	'Yes'- visited	376
2	'No'- not visited	124
	Total persons contacted	500

Table: Example of frequency table [classification of responses of respondents]

If the respondents are to be classified based on the age, the frequency is presented as in table 8.6

Sl.No	Classification (Age group)	Frequency (no. of respondents in each class)	Percentage
1	<20 years	84	16.8
2	21-40 years	302	60.4
3	>40 years	114	22.8
	Total	500	100

Table: Example of frequency table [classification of respondents based on age]

Cross-tabulation and contingency table:

It is organizing data by groups, categories or classes to facilitate comparisons. It is a joint frequency distributions of observations on two or more sets of variables.

If the respondents in the above example are classified based on the gender and their response it becomes a 2x2 contingency table

Gender	Frequency (visit to Brookfields)		Total
	Yes	No	
Male	361	31	392
Female	15	93	108
Total	376	124	500

Table Example of 2x2 cross table/contingency table

If the respondents in the same example are classified on gender and 3 levels of qualification it becomes a 2x3 contingency table)

Sl. No.	Gender	Qualifications	Total

		Matriculates	Graduates	Post graduates	
1	Male	30	275	87	392
2	Female	20	80	8	108
	Total	50	355	95	500

Table: Example of classification of the respondents based on gender and qualification.

Construction of Tables: Guidelines

- All the tables must be numbered for easy reference. It could be in a consecutive serial order such as Table1, Table 2, Table 3 and etc or chapter wise as 1.1., 1.2, 1.3 … for the tables in chapter 1 and 2.1, 2.2, 2.3… for the tables in chapter 2.
- Every table should have a title or caption which is clear and concise to give a clear description of the contents
- The units of measurement must always be indicated
- If any symbols or abbreviations are used due to lack of space, they must be identified in the footnotes below the table.
- If the data presented are not original, the sources of data must be indicated below the table.
- There should be proper alignment of column, figures, decimal points, plus or minus signs and etc.
- Miscellaneous and exceptional items are placed at the last row (s)
- While discussing, tables are referred to only by number and not 'as table given above' or 'the following table'.
- Tables wider than the size of the report are placed side wise so that the title is near the margin or binding side.
- Tables should be made as logical, clear, concise and simple as possible.

Pictorial presentation

Diagrams are among the most frequently used methods of presenting quantitative data. Diagrams are relatively easy to understand and interpret. While bar chart and pie chart are used for nominal or ordinal variables, histogram is used for interval/ratio variables.

Example: A survey was conducted among 180 people of varying age group 11 to 60 years to know why people go to gym. The sample survey showed that 18 people visit gym just to relax and spend time, 62 to maintain fitness, 66 to lose weight and 34 to build strength. The survey also showed that among the 180 members who visit gym regularly 6 are in the age group of 11-20 years, 78 are 21-30 years, 46 are 31-40 years, 42 are 41-50 years and 8 are 51-60 years.

Histogram: A histogram is a diagrammatic presentation of the frequency distribution of a continuous variable. In a histogram the horizontal axis represents the class boundaries and the vertical axis is equal or proportional to the frequencies, so long as all the classes are of the same width. Histograms are used when it is possible to group the variables' values into intervals. A histogram is a series of rectangles each proportional in width to the range of values within a class and proportional in height to the no of items falling in the class. They are constructed with bars that represent data values where each value occupies an equal amount of area within the enclosed data. If the classes in the distribution are of equal width then the vertical bars in the histogram are also of equal width. The height of the bar for each class corresponds to the no of items in the class.

The data given in the above example are presented in a histogram as given in figure

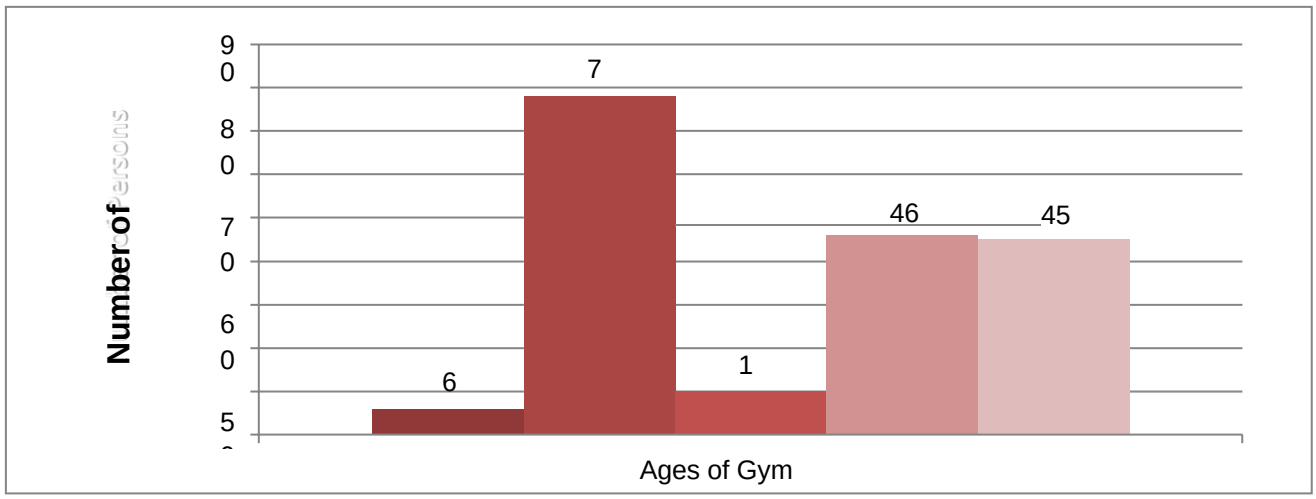

Figure Example of histogram showing ages of gym visitors

Bar Chart

In bar charts each bar represents the no. of people or variable in each category. The bars could be vertical or horizontal.

The people who visit gym (in the said example) for various reasons are given as bar chart in figure .

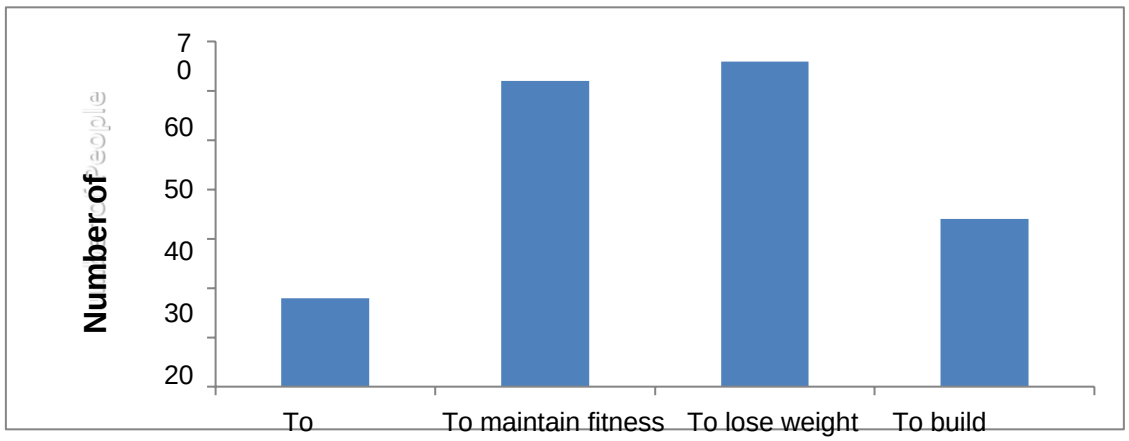

Figure: Example of bar chart

Pie Chart

- In pie chart percentages of subjects (or variables) are shown. The percentage that each slice representing whole sample is given in the pie chart
- Pie chart must show 100% of the subjects being graphed
- To avoid separate legends, the slices are labelled with call-outs with percentage
- The largest slice is placed at the twelve O'clock position and moved clockwise in descending order

The percentage of people who visit gym for various reasons (in the given example) are shown as pie-chart in figure.

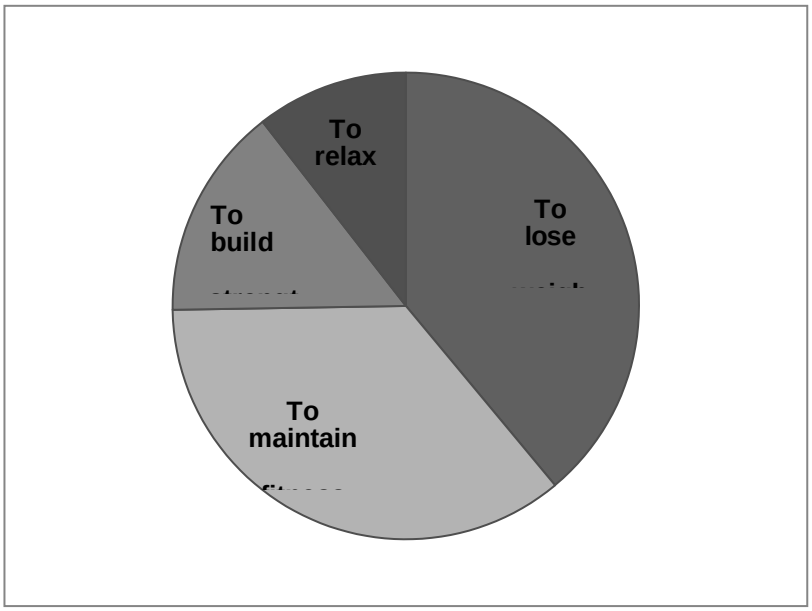

Figure: Example of pie chart (why people go to gym in a given example)

The Contents in this E-Material has been taken from the text and reference book as given in the syllabus