Dr. Prakash R. Rodiya

Dept. of Commerce

Research Methodology

4. DATA COLLECTION, DATA PROCESSING AND DATA ANALYSIS

Definition

" Data are facts, figures and other relevant materials past and present serving as bases for study and analysis".

Meaning of Data

The search for answers to research questions calls collection of Data. "Data are facts, figures and other relevant materials, past and present, serving as bases for study and analysis".

Types of Data

The Data needed for social science Research may be broadly classified into:

a) Data pertaining to human beings

b) Data relating to Organisations

c) Data pertaining to territorial area.

A) Personal Data (relating to Human beings) are of two types.

a) Demographic and socio-economic characteristics of individuals. Like name, sex, race, social class, relation, education, occupation, income etc.

b) Behavioural Variables: Attitudes, opinion knowledge, practice, intensions etc.

B) Organistion Data:- Consist of data relating to an organizations, origin ownership, function, performance etc.

C) Territorial Data :- are related to geo-physical characteristic, population, infrastructure etc of divisions like villages, cities, taluks, distinct, state etc.

Importance of Data

The data serve as the bases or raw materials for analysis without Data no specific inferences can be drawn on our study. The reliability of data determines the quality of research. Data form the basis of testing hypothesis data provides the facts and figures for constructing measuring scale. The scientific process of research can be carried out only through accurate data.

Sources of Data

The sources of Data may be classified into a) primary sources b) Secondary sources.

a) Primary Sources.

Primary sources are original sources from which the researcher directly collects data that have not been previously collected. Primary Data are first –hand information collected through various methods such as observation, interview etc.

b) Secondary Sources

These are sources containing data which have been collected and compiled for another purpose. The secondary sources consist of readily available and compiled statistical statements and reports. Secondary sources consist of not only published but also unpublished records. They consist of Data over which a researcher has no original control.

Collection of Primary Data

The important methods of Primary data are:

1. Observation
2. Interviewing
3. Schedules
4. Questionnaire

Observation

I. Observation means viewing or seeing.

We observe things while we are awake.

Observation is classical method of scientific enquiry

Definition

Observation may be defined as "a systematic viewing of a specific phenomenon in its proper setting for the specific purpose of gathering data for a particular study". Observation as a method includes both seeing and hearing"

Characteristics of Observation

1. It is both physical and mental activity. The observing eyes catches many things which

are sighted but attention is focused on that data are pertinent to a given study.

2. Observation is selective a researches does not observe anything and everything but select the range of things to be observed on the basis of the nature, scope and objectives of his study.
3. Observation is purposive and not casual. It is made for the specific purpose of not in things relevant to the study.
4. Observation captures to the study. Social context in which a persons behaviour occurs.
5. Observation is based on standarised tools.

Types of Observation

With reference to the investigators role observation is classified into:

a) Participants observation
b) Non-Participates Observation

In terms of mode of observation. Observation is classified into

a) Direct Observation
b) Indirect Observation

With reference to the system adopted observation is classified into
Controlled observation and uncontrolled observation

a) Participant Observation

In this observation, the observer is a part of the phenomena on group which is observed and he as both a observer and participants.

It has certain advantages

1. The Observer can understand the emotional reactions of the observed group.
2. The observer will be able to record the context which gives meaning to the observed behaviour

The main Disadvantages are:

1. The Participant observer narrows his range of observation.
2. In this type of observation , The objectivity is lost.
3. The clear demand makes in accuracy in recording.

Non Participant observation

In this method the observer stands apart and does not participate in the phenomenon observed. Naturally there is no emotional involvement on the part of the observer. This method calls for skill in recording observation in an unnoticed manner.

D) Direct Observation

This means observation of an event personally by the observer when it takes place. This is a flexible method.

Indirect Observation

This does not involve the physical presence of the observer and the recording is done by mechanical autographic or electronic devices.

E) Controlled Observations – Maximum control over extrinsic and intrinsic variable

F) Uncontrolled Observations – No control over extrinsic and intrinsic variable.

Planning of Observation

The use of observation method requires proper planning; first, the researcher should carefully examine the relevance of observational method.

Second, he must identify the specific investigative questions.

Third, he must decide the observation content .

Four, observer should be selected and trained

Tools of observation

Systematic observation requires certain devices. They are:

Schedules, diary, Cards, Cameras, tape recorder, video tape, tape, Barometers, screen and mirrors, closed circuit television, pocket calculators etc.

Advantages of observations

Observation has certain advantages

1. The main advantage is that it makes it possible to study behaviour as it occurs.
2. Data collected by observation may describe the observed phenomena as they occur in their natural setting
3. Observation is more suitable for studying subjects who are unable to articulate eg. Children, animals, birds etc.
4. Observation makes it possible to capture the whole events.
5. It is easier to make disguised observation
6. Observation for mechanical devices it can be recorded correctly.

Limitations

1. Observation is of no use in studying past events.
2. It is unable to study opinion and attitudes.
3. The researcher has to wait for the events
4. Observation is very expensive.

II. Interviewing

Interview is one of the major methods of data collection. It is often superior to other data collection methods. People are usually more willing to talk than to write.

Definition

"It may be defined as a two way systematic conversation b/w an investigator and an informant initiated for obtaining information relevant to a specific study"

Types of Interview

a) <u>Structured or directive Interview</u>

This is an Interview made with a detailed standard schedule. The same questions are put to all the respondents and the same order. Each question is asked in the same way in each interview. This type of Interview is used for large scale formulated surveys.

It has certain advantages

1. Data from one Interview to the next can be easily compared
2. Recording and coding of data do not cause any problem.

It has some limitations

1. It tends to loss spontaneity of conversation
2. The respondants view is minimized.

b) Unstructured or Non-directive Interviews

Respondent to talk freely about a given topic and a pre-planned schedule is not used.

It has some advantages

1. It is similar to natural conversation
2. It provides great opportunity to explore a problem.

It has some limitations

1. The data is not comparable
2. Time is wasted for unproductive conversation

c) Focused Interview

This is a semi- structural interview the investigator attempts to focus the discussion on the actual effect of a given experience to which the respondents opinion, emotions or conditions on the basis of an interview guide. This required training and skilled.

Interview Processing

The Interviewing process consist of the following stages

1. Preparation
2. Introduction
3. Developing rapport
4. Carrying the interview forward

5. Recording the Interview
6. Closing the Interview

Interview problems

In a personal Interview there are certain problems

I. Response Problem

There are different types of response problem

a) Non-Response

Non-response refers to failure to obtain response from respondents here respondent remains silent or refuse to answer

b) Partial response

In partial response the respondent give an incomplete answer

c) Irrelevant response

In irrelevant response respondent gives totally irrelevant answers.

II. Interviewers bias

The Interviewer may resort to cheat by taking up data without actually interviewing he may use manipulations by rephrasing the question etc.

III. Non availability

Another major problem of Interviews is the non-availability of respondent. A respondent may be a too busy or out of stations.

Features of Interview

1. Interview is not a casual conversation but a conversation with a specific purpose.
2. The participant, the Interviewer and the respondent both are strangers.
3. Interview is a mode of obtaining verbal answers to verbal questions.
4. Interview is an interaction process.
5. Interview is not a chemical examination.

TELEPHONE INTERVIEW

Telephone Interview is a non-personal method of data collection. It may be used as a major method or supplementary method of data collection. This method is used when the universe is very large and the survey is will completed to a limited period of time and the

universe is widely scatted. It has some advantages.

1. It is a low cost method
2. Large No. of respondents can be covered within a small time and it does not require field work.

It has some limitations:-

1. There is a time limit to the telephone conversation.

GROUP INTERVIEW

A Group Interview may be defined as a method of collecting primary data in which a No. of individuals with a common interest interact with each other a group may consist of six to eight individuals. The interviewer acts as the discussion leader. Information is obtained self administered questions. It is a popular method and has the following advantages.

1. The respondents comment freely and detailed.
2. This method is highly flexible.
3. They didn't have watch the interview as observed.

It has some limitations also

5. It is difficult to representative samples.

There is the possibility of one individual dominating others.

IV. Schedules and Questionnaire

Schedules and Questionnaires are the most common instruments of data collection. These two types of tools have much in common. Both of them contain a set of questions logically related to a problem under study; Both aim at collecting responses from the respondents. But both are different. A schedule is used as a tool for interviewing, a questionnaire is used for mailing. The schedule is filled by the interviewer in a face-to-face interview; whereas a questionnaire is filled by the respondent himself.

V. QUESTIONNAIRE

This method of data collection is quite popular particularly in case of big enquiries. It is being adopted by private individuals, research workers and even by governments. In this method a questionnaire is sent (usually by post) to the persons concerned with a request to answer the questions and return the questionnaire.

A Questionnaire contains a number of questions printed or typed in a definite order.

Then it is mailed to the respondents who are expected to read and answer to questions and return by writing down the answer. The respondents have to answer the questions on their own. It has the following merits.

1. There is a low cost even when the universe is large and it is widely spread geographically
2. It is free from the bias of the interviewer answers are in respondents own words.
3. The respondents have enough time to read and answer the questions.
4. Respondents who are not easily approachable can also be reached conveniently.
5. Large samples can be used; so the method is dependable and reliable.

Limitations

The main limitations of this system are;

1. Low rate of return of duly filled questionnaire
2. It is possible only in case of educated respondents.
3. The control over the questions is lost once it is sent.
4. There is inflexibility of questions.
5. There is possibility of unclear answers.
6. This method is a slowest one.

SCHEDULES

This method of data collection is very much like the collection of data through questionnaire with little difference which likes in the fact that schedules are being filled in by enumerators who are specially appointed for this purpose. These enumerators along with schedules go to respondents, put to them the questions from the performer in the order questions are listed and record the replay in the space meant for the same Performa. This method requires the selection and training of enumerators to fill up the schedules and they should be carefully selected. Enumerators should be intelligent and must be able to find out the truth. The enumerators should be honest sincere and hard working. This method is very useful because it yield good results. Population censes all over the world is conducted through this method.

Differences between Schedule and Questionnaire

1. The Questionnaire is generally sent through mail to informants. The schedule is generally filled by the research worker.
2. To collect data through questionnaire is relatively cheap. To collect data through

schedule is relatively more expensive.

3. Non- response is high in case of questionnaire whereas in schedule response is very high.
4. In Questionnaire there is no personal conducts. But in a schedule there is a face-to-face contact.
5. The questionnaire method is used only when respondents are literate.
6. Along with schedules observation methods can be also used.

Types of Questions

a) Structural Question

Questions which allow only a few alternative ways of answering are structural questions. The simplest example of a structural question "What is your age"? there can be only one answer to this question.

b) Dichotomous Questions (two choice questions)

Dichotomous Questions result in Yes or No answers. For eg..Are you a member of lions club?

c) Multiple choice questions

Here the answer is selected from among several alternatives

d) Ranking questions

Here different alternative answers are given for question and the respondent is asked to rank them or show his preference by number income 1,2,3, etc.

Questions not to be asked

1. Vague questions should be avoided
2. Leading questions should be avoided.
3. Presuming questions should not be included.
4. Hypothetical questions should not be included.

DATA PROCESSING AND DATA ANALYSIS

Data Processing

Data processing is an intermediately stage of work b/w data collection and data

analysis. The complete Instruments of data collection that is Interview schedule Questionnaire, data sheet etc contain a west mess of data. They cannot straight way provide answers to research question. They are like raw materials and they need processing. The following are the major steps in data processing.

1. Editing

The 1st step in processing of data is editing of data. Editing is a process of checking to detect and correct errors and concessions. Editing is done at 2 stages.

a) Field Editing

During the time of Interviewing. The interviewer cannot always record response completely and legibly. Therefore after each Interview is over the researcher should review the schedule to complete abbreviated responses, rewrite eligible response and correct omissions.

b) Office Editing

All completed schedules and questionnaire should be thoroughly checked in the office for completeness accuracy and uniformity.

2. Classification and Coding

a) Classification

The edited data are classified and coded. The responses are classified into meaningful category, so as to bring out there essential pattern. By this method several ended responses are reduced into 5 or 6 appropriate categories containing critical information needed for analysis. Suppose the responses to a question an occupation in a survey consist of items as business executive, share broker, electricians, Driver, farm labourer, lawyer, college teacher, medical practitioner barber, carpenter, accountant and gold smith.

This data are not amenable for analysis. So they can be classified as

1. Professional and managerial Business executive, college teacher, lawyer, medical practitioner.
2. Clerical: Accountant, Assistant
3. Skilled labours: Share broker, electricians, carpenter, barber, Gold smith
4. Unskilled labourer: Farm labourer

b) Coding

Coding means assigning numerals or other symbols to the category of responses for each question. A coding scheme is designed on the basis of the concerned category. The coding schemes with the assigned symbols together with specific coding instructions may be assembled in a book. The codebook will identity a specific items of variable or observation and code number assigned to each category of that item.

Occupation		**Religion**	
Salaried	- 1	Hindu	- 1
Profssional	- 2	Christian	- 2
Business	- 3	Muslim	- 3
Retired	- 4	Budhist	- 4
House Wife	- 5	Jainsist	- 5

3. Transcription

When only a few schedules are processed and hand tabulated, Tabulation can directly we made from the schedules. On the other hand direct tabulation from the edited schedule or questionnaire is difficult if the number of schedules and the number of responses in them are large.

Suppose an Interview schedule contain 180 responses requiry tabulation and 210 simple and cross tables are to be constructed, each schedule has to be handled at least 210 for tabulation. This will result in mutilation of the schedule. In order to avoid this draw able data contained in schedule or questionnaire are transferred to another material for tabulation. This intermediately process is called transcription. There are two methods of transcription Manual or mechanical

4. Tabulation

After the transcription of data is over data are summarised and arranged in a compact form for further analysis. Thus tabulation is the process of summonsing raw-data and discipline them of compact statistical tables. Tabulation can be done by hand or by mechanical or electronic devices.

5. Graphic Representation

In presenting the data of frequency distinction and statistical computation, It is often desirable to use appropriate to forms of graphic representation. In addition to tabular forms. Graphic representation involves the use of graphic charts and other pictorial devices reduced large masses of statistical data to a form that can be quickly understood at a glance. The devise of graphic representation are useful for non-technical people or general public. Graphic representation must be planned with almost care. Graphic forms used must be simple clear and accurate. The most commonly used graphical forms are

a) Line graph or charts
b) Bar charts
c) Segmental representations
d) Pictograph