

# LECTURE 13 SOURCES OF DATA

DR GAURAV SANKALP

## SOURCES OF DATA

The sources of data may be classified into a) primary sources b) secondary sources. Both the sources of information have their merits and demerits. The selection of a particular source depends upon the (a) purpose and scope of enquiry, (b) availability of time, (c) availability of finance, (d) accuracy required, (e) statistical tools to be used, (f) sources of information (data), and (g) method of data collection.

## PRIMARY SOURCES OF DATA

Primary sources are original sources from which the researcher directly collects data that have not been previously collected e.g., collection of data directly by the researcher on brand awareness, brand preference, brand loyalty and other aspects of consumer behaviour from a sample of consumers by interviewing them. Primary data are firsthand information collected through various methods such as observation, interviewing, mailing etc.

According to **P. V. Young**, “primary sources are those data gathered at first hand and the responsibility so of their compilation and promulgations remaining under the same authority that originally gathered them.”

In the words of **Watter R. Borg**, “Primary sources are direct describing occurrences by an individual who actually observed on witness for occurrences.”

## ADVANTAGE OF PRIMARY DATA

- It is original source of data
- It is possible to capture the changes occurring in the course of time.
- It flexible to the advantage of researcher.
- Researchers know its accuracy.
- Only that data are collected which meet outs the objective of research project.
- In maximum methods of primary data collection researchers know who the respondents are so face to face communication is there.
- It is most authentic since the information is not filtered or tampered.
- Extensive research study is based of primary data

## DISADVANTAGE OF PRIMARY DATA

- Primary data is expensive to obtain
- It is time consuming
- It requires extensive research personnel who are skilled.
- It is difficult to administer.
- Chances of biasness are at great extent.
- Biasness can also be there on the part of respondent. Wrong answer can be given by then which may affect the accuracy of data.
- It may have narrow coverage. It means researchers may collect data only within his/her reach or according to his mindset.

## METHODS OF COLLECTING PRIMARY DATA

The following are the methods of collecting Primary Data:

1. PERSONAL INTERVIEW
2. OBSERVATION METHOD
3. DATA COLLECTED THROUGH MAIL
4. WEB METHOD
5. TELEPHONE METHOD

1. **PERSONAL INTERVIEW:** Under this method the researcher personally visits the area of enquiry, establishes personal contact with the respondent and collects necessary facts and figures.

*Example:* If a researcher wants to know about the family income of persons in a particular area, he goes personally to the area and collects data on the basis of personal contacts. It will be direct personal interview.

### Advantages:

The main advantages of personal interview are:

- (i) **Generally yields highest cooperation and lowest refusal rates:** Since personal visits are made to the respondents, to the refusal rates are low as better explanation about the research work can be given to the respondent.
- (ii) **Allows for longer, more complex interviews:** Since contact is made personally and the researcher has time to explain the about the complexity of the research, in

details, to the respondent, so he may go for complex interviews where questions have to be framed at the point of the respondent.

- (iii.) **High response quality:** The response quality of the respondent can be judged at the time of collecting the data, so the quality of data is controlled.
- (iv.) **Multi-method data collection:** Under this method the person collecting the data may change the method of data collection if he is not getting appropriate data by from his existing method.

### **Disadvantages**

The main disadvantages of personal interview are:

- (i.) **Most costly mode of administration:** Since personal visits are made for conducting the interview and also two respondents may be at far away distance, so the cost of collecting the data increases, as it will include the transportation cost of the person collecting the data.
- (ii.) **Longer data collection period:** Since the data is collected by personal visit, so it becomes a very slow process to collect data. It" takes comparative longer period to collect the data under this method.
- (iii.) **Interviewer concerns:** Since the interview is conducted by making personal visits, so there is always a chance, that the respondent may not respond to the process, or may give his own suggestion for changes in the research and provide data according to the changes.

**2. OBSERVATION METHOD:** Under this method, the researcher collects information directly through systematic watching and noting the phenomena as they occur in nature with regard to cause and effect or mutual relations rather than through the reports of others. It is a process of recording relevant information without asking anyone specific questions and in some cases, even without the knowledge of the respondents.

### **Advantages:**

- (i.) **The respondent will provide data:** In any case the data is collected from the respondent by observing the response pattern and the respondents are unable or reluctant to provide information.

- (ii.) **Data Accuracy:** The method provides deeper insights into the problem and generally the data is accurate and quicker to process. Therefore, this is useful for intensive study rather than extensive study.

### **Disadvantages**

- (i.) Unable to predict the occurrence of data: In many situations, the researcher cannot predict when the events will occur. So when an event occurs there may not be a ready observer to observe the event.
- (ii.) No true data: As the respondent may be aware of the observer and as a result may alter their behavioural pattern.
- (iii.) Paradigm: Since this method solely depends on the observation power of the researcher, so due to lack of training and paradigm the researcher may not observe the things as they occur.
- (iv.) Not suitable for large research: This method cannot be used extensively if the inquiry is large and spread over a wide area.

**3. DATA COLLECTED THROUGH MAIL:** Under this method the data is collected by sending letters to the respondent. A letter may contain questionnaire and the respondent is required to respond back.

### **Advantages**

- (i.) **Generally lowest cost:** As compared to other form of data collection, this method is cheapest as the questionnaire maximum of one page. Further to cut the cost, the researcher may go through print media to get its questionnaire distributed to the respondent.
- (ii.) **Can be administered by smaller team of people:** This research can be administered with few staffs as only the office staffs are required and no field staffs are necessary.
- (iii.) **Access to otherwise difficult to locate, busy populations:** As the research is done through mail, so the researcher could get the data from the respondent otherwise it is difficult to locate amongst the busy population.
- (iv.) **Respondents can look up information or consult with others:** As the respondent don't have to respond to the queries of the researcher, so the

respondent gets enough time to understand the information required and he may even consult other person to provide the accurate data.

### Disadvantages

- (i.) **Most difficult to obtain cooperation:** As the respondent has the option to respond back or not, so the researcher may find it difficult to collect information as all of its questionnaire sent may not return with the data. Also there is problem of delivery of mails to the respondent.
- (ii.) **No researcher involved in collection of data:** As all the work of collection of data are through mails, so there are no researcher involved in the process, so the respondent may sometimes find it difficult to understand the queries raised by the researcher.
- (iii.) **Need good sample:** As the mails are sent with the help of some database, so the researcher may not know, how the respondent will react? The respondent may not provide appropriate data as required by the researcher.
- (iv.) **More likely to need an incentive for respondent:** In order to make respondent give response to the letter sent, there should be some incentive scheme to be attached with the mail, otherwise, the response rate will be very poor.
- (v.) **Slower data collection period:** As the time needed for delivery of mails to the respondent and back, so this method is much slower than how the data is collected through telephone or personal interview or web mail.

4. **WEB METHOD:** Under this method, the data is collected through internet. This method can be further divided into two groups

- A. **Through Polling:** The researcher may put the information on a web server and the respondent may require responding to the information through online poll, or blog.
- B. **Through Mails:** The researcher may also opt to send emails to various respondents and may give them the option to respond back.

### Advantages

- (i.) **Lower cost:** As the research work is completed online and the data are collected in the database, so the researcher may not require any paper, postage, mailing, data entry costs.

- (ii.) **Can reach international populations:** As the research work is done online, so the researcher may also involve international respondent for collection of data at no extra cost or effort.
- (iii.) **Time required for implementation reduced:** As the respondent is required to send the response online, and the data is also collected, so the researcher don't have to waste time in compiling the data and interpreting the data. The researcher may directly go for interpretation.
- (iv.) **Complex patterns can be programmed:** As the research work is online, so the researcher may go for complex research activity as all the queries of the respondent are immediately handled by the researcher.
- (v.) **Sample size can be greater:** As the research work is online, the sample size for the research may be greater as the population outside the country can be included.

#### Disadvantages

- (i.) **Limitation of technology:** As in India, approximately 55% of homes own a computer; 30% have home e-mail, so the choice of population is restricted for the researcher.
- (ii.) **Representative samples difficult:** Since the access to the technology is difficult for general population, so the data collection activity cannot generate random samples of the population.
- (iii.) **Differences in capabilities of people's computers and software for accessing Web surveys:** Since each person has different capabilities and knowledge about the usage and utility of the web, so a good respondent may not provide the sample for the research.
- (iv.) **Difference in people's response:** The researcher cannot ascertain, if the same person has given response to the survey. If a research activity is performed by some other person, then the quality of response will be different.
- (v.) **Different ISPs/line speeds limits extent of graphics that can be used:** If the researcher is using graphic display to explain the theme or complexities of his research to the respondent, so it is quite possible that some of the respondent may not get the graphics displayed on their computer due to ISPs/line speeds limit or restriction of usage. In' such case the data collected will not be accurate.

**5. TELEPHONE METHOD:** Under this method, the researcher calls the respondent and collects the data over the telephone. He may use the telephone numbers available on the Telephone directory, and select the samples from the given population.

#### **Advantages**

- (i.) **Less expensive than personal interviews:** As the research can be completed over the phone, so this is less expensive than the personal interview and that the data are collected quickly.
- (ii.) **Samples from general population:** As telephones are accessible to the general population, than the Web Method, so more samples can be collected from larger population.
- (iii.) **Shorter data collection period than personal interviews:** As limited tools can be used to explain the research objective to the respondent, so it takes much lesser time to collect data. Also the larger population can be reached in very short time.
- (iv.) **Researcher administration:** As the researcher can explain and listen to the queries of the respondent, so in this method there is direct control of the researcher over the subject for data collection than the mails or web mails.
- (v.) **Better control and supervision of Researcher:** Similarly, as the researcher is contacting the respondent from his place, so the researcher may refer any literature during the process of collecting data. This is restricted in case of personal interview.
- (vi.) **Better response rate than mail for list samples:** For Comparative Scaling techniques this method is easy and effective as the researcher will provide the list and respondent have to choose from the available lists.

#### **Disadvantages:**

- (i.) **Biased against households without telephones, unlisted numbers:** As still most homes in India, don't have telephone numbers, or their numbers are not listed in the telephone directories due to having pre-paid connections. In such case such persons don't get equal opportunities to appear for research activity.
- (ii.) **No response:** If the respondents don't pick up the telephone on time or if the call is missed, in such case the researcher may select any other person as sample.
- (iii.) **Questionnaire constraints:** Complex questionnaire prepared by the researcher cannot be used in this method, if next question of the questionnaire depends on the

answer of the respondent then such types of questionnaire cannot be included in the questionnaire.

- (iv.) **Difficult to administer questionnaires on sensitive or complex topics:** Since the researcher is getting the names of the samples from the telephone directory, and he is not aware about the economic, social or emotional situation of the samples, so it is very difficult to administer questionnaires on sensitive or complex topics.

## **SECONDARY SOURCES OF DATA**

These are sources containing data which have been collected and compiled for another purpose. The secondary sources consists of readily compendia and already compiled statistical statements and reports whose data may be used by researchers for their studies e.g., census reports , annual reports and financial statements of companies, Statistical statement, Reports of Government Departments, Annual reports of currency and finance published by the Reserve Bank of India, Statistical statements relating to Cooperatives and Regional Banks, published by the NABARD, Reports of the National sample survey Organization, Reports of trade associations, publications of international organizations such as UNO, IMF, World Bank, ILO, WHO, etc., Trade and Financial journals newspapers etc.

Secondary sources consist of not only published records and reports, but also unpublished records. The latter category includes various records and registers maintained by the firms and organizations, e.g., accounting and financial records, personnel records, register of members, minutes of meetings, inventory records etc.

### **FEATURES OF SECONDARY DATA**

Though secondary sources are diverse and consist of all sorts of materials, they have certain common characteristics.

First, they are readymade and readily available, and do not require the trouble of constructing tools and administering them.

Second, they consist of data which a researcher has no original control over collection and classification. Both the form and the content of secondary sources are shaped by others. Clearly, this is a feature which can limit the research value of secondary sources.

Finally, secondary sources are not limited in time and space. That is, the researcher using them need not have been present when and where they were gathered.



## **USES OF SECONDARY DATA**

The second data may be used in three ways by a researcher. First, some specific information from secondary sources may be used for reference purpose. For example, the general statistical information in the number of co-operative credit societies in the country, their coverage of villages, their capital structure, volume of business etc., may be taken from published reports and quoted as background information in a study on the evaluation of performance of cooperative credit societies in a selected district/state.

Second, secondary data may be used as bench marks against which the findings of research may be tested, e.g., the findings of a local or regional survey may be compared with the national averages; the performance indicators of a particular bank may be tested against the corresponding indicators of the banking industry as a whole; and so on.

Finally, secondary data may be used as the sole source of information for a research project. Such studies as securities Market Behaviour, Financial Analysis of companies, Trade in credit allocation in commercial banks, sociological studies on crimes, historical studies, and the like, depend primarily on secondary data. Year books, statistical reports of government departments, report of public organizations of Bureau of Public Enterprises, Censes Reports etc, serve as major data sources for such research studies.

## **ADVANTAGES OF SECONDARY DATA**

Secondary sources have some advantages:

1. Secondary data, if available can be secured quickly and cheaply. Once their source of documents and reports are located, collection of data is just matter of desk work. Even the tediousness of copying the data from the source can now be avoided, thanks to Xeroxing facilities.
2. Wider geographical area and longer reference period may be covered without much cost. Thus, the use of secondary data extends the researcher's space and time reach.
3. The use of secondary data broadens the data base from which scientific generalizations can be made.
4. Environmental and cultural settings are required for the study.
5. The use of secondary data enables a researcher to verify the findings bases on primary data. It readily meets the need for additional empirical support. The researcher need not wait the time when additional primary data can be collected.

## **DISADVANTAGES OF SECONDARY DATA**

Although secondary data are easy to access and cost-effective, they also have significant limitations:

1. The secondary data are not up-to-date and become obsolete when they appear in print, because of time lag in producing them. For example, population census data are published two or three years later after compilation and no new figures will be available for another ten years.
2. Data may be too broad-based that is, not specific enough to adequately address the firm's research questions.
3. The units in which the data are presented may not be meaningful.
4. The source of the data may not provide sufficient supporting material to allow the researcher to judge the quality of the research.
5. The data sources may lack reliability and credibility. Some secondary data may simply be inaccurate.
6. The most important limitation is the available data may not meet our specific needs. The definitions adopted by those who collected those data may be different; units of measure may not match; and time periods may also be different.
7. The available data may not be as accurate as desired. To assess their accuracy we need to know how the data were collected.
8. Finally, information about the whereabouts of sources may not be available to all social scientists. Even if the location of the source is known, the accessibility depends primarily on proximity. For example, most of the unpublished official records and compilations are located in the capital city, and they are not within the easy reach of researchers based in far off places.

## **METHODS OF COLLECTING SECONDARY DATA**

The researcher may get Secondary data from two sources (a) Published; (b) Unpublished. While we are going to discuss about published data, the unpublished data may be in records of Government or private organizations, research organizations, research scholars etc. However, these data being unpublished is not freely available and the details about the sources remain with few persons.

Most of the researcher use published data as they are available from the following sources:

1. **Newspaper and Magazines:** Statistical data on a number of current socioeconomic subjects can be obtained from data collected and published by some reputed newspapers, magazines, periodicals, etc.
2. **Published Articles of an individual:** Many times some research studies are carried on at individual level and are published in magazines in the form of articles or in the form of books. Though these studies are based on research works of limited area, however they may provide useful information for further research.
3. **Government Publication and Gazetteer:** Various ministries and departments of Central Government and State Governments publish data regularly on a number of subjects. These data are considered reasonably reliable for research work.
4. **Reports from Commissions and Committees:** The government constitutes committees and commissions for the study and enquiry of various problems from time to time, which submit their reports after collection and analysis of required data and information. The information contained in such reports is very useful and reliable.
5. **Publications from various Organizations:** Many important universities and semi-government research organizations also publish their research studies and these publications are important sources for analytical statistical information.
6. **Private Data Publication:** Recent trends have emerged where the private organizations are collecting and compiling various data for further usage as research data.

#### 4.5 DIFFERENCE BETWEEN PRIMARY DATA AND SECONDARY DATA

- (a). Primary data are collected by the researcher himself, although the secondary data has been collected previously by other researcher.
- (b). Primary data are collected and used first time. However, on the secondary data, some decisions has been made previously, such decisions may or may not be useful for the researcher now.
- (c). Since primary data is collected by the researcher himself, so it relates directly to the research objective *or* may be more close to research objective. How many parts of secondary data need not to be related to the research objective?
- (d). Since primary data is collected by the researcher, so it is more time taking activity for the researcher than to get the secondary data.

The students should note here that the primary data and secondary data differ in the degree of impact only and that same set of data may be secondary in the hands of one and primary in the hands of another. As Secrist has said "the distinction between primary and secondary data is largely one of degree. Data which are secondary in the hands of one party may be primary in the hands of another" Example, if population statistics are primary which it is collected under population census but it becomes secondary when it is used for some researcher.

RM MOOC UPRTOU