1. Introduction

Welcome to the series of E-learning modules on undertaking of sample surveys. In this module, we are going to cover the principal steps of a sample survey, methods and tools used for data collection and principal findings from the sample survey.

By the end of this session, you will be able to:

- Explain sample design or survey
- Explain the principal steps of a sample survey
- Explain the methods and tools used for collection of information
- Explain the Statistical Analysis or Principal findings

A sample design is a definite plan for obtaining a sample from a given population. It refers to the technique or the procedure, the researcher would adopt in selecting items for the sample.

A sample design may also lay down the number of items to be included in the sample. Sample design is determined before the data are collected.

There are many sample designs which a researcher can choose. Some designs are relatively more precise and easier to adopt than others are. Researcher must select or prepare a design, which should be reliable and appropriate for his research study.

While undertaking a sample survey one must pay attention to the following points.

Principal Steps of a Sample Survey

1. Objectives of the survey:

The first step in any survey is, deciding what you want to learn. The goals of the project determine whom you will survey and what you will ask them. If your goals are unclear, the results will probably be unclear.

Some typical goals include learning more about:

- The potential market for a new product or service
- Ratings of current products or services
- Employee attitudes
- Customer or patient satisfaction levels
- Reader or viewer or listener opinions
- Association member opinions
- Opinions about political candidates or issues
- Corporate images

These sample goals represent general areas. The more specific you can make your goals, the easier it will be to get usable answers.

Hence, the first step is to define in clear words about the objectives of the survey. Sometimes, the sponsoring agents also will not know about the objectives of the survey like, what it wants and how it is going to make use of the results.

However, they should take utmost care so that their objectives should go along with the available resource in terms of money, manpower and the time limit required for the completion of the survey.

2. Definition of the population to be sampled:

The main step in developing any sampling design is to clearly define the set of objects from which the sample is chosen technically called as a Universe or Population.

For example, in sampling of farms, clear rules must be framed regarding the size, shape etc. of the farm giving importance for the border line cases, so that an investigator will be able to decide in the field without any hesitation whether to include or not to include a given farm in the population.

3. The Frame and the Sampling Units:

The population under study must be able to divide into some parts called as sampling units for the selection of the sample. Sampling units must cover the whole population and they must be distinct and non overlapping, that is, every unit of the population must belong to one and only one sampling unit.

Hence, a decision should be taken concerning a sampling unit before selecting a sample. Sampling unit may be a geographical one, such as state, district, village etc. or a construction of a unit such as house, flat etc. or it may be a social unit such as family, club, school etc. or it may be an individual.

2. Frame

Frame can be defined as some list or map or other acceptable material which covers the population decided upon and which serves as a guide for the population to be covered. Since frame is the one, which determines the structure of the sample survey, it has become one of the major practical problems.

A frame which has been already prepared for some other purpose has to be scrutinized and should be checked to see that it is free from all sort of defect like unknown amount of duplication and should be brought up-to-date before using them. A good experience helps in constructing a good frame.
A Frame should be comprehensive, correct, reliable and appropriate. It is extremely important for the frame to be as representative of the population as possible.

4. Data to be Collected:
There are two main components in determining whom you will interview. The first is to decide what kind of people to interview. Researchers often call this group the target population. If you conduct an employee attitude survey or an association membership survey, the population is obvious.

If you are trying to determine the likely success of a product, the target population may be less obvious. Correctly determining the target population is critical. If you do not interview the right kinds of people, you will not successfully meet your goals.

The next thing to decide is how many people you need to interview. Statisticians know that, a small representative sample will reflect the group from which it is drawn. The larger the sample, the more precisely it reflects the target group.

However, the rate of improvement in the precision decreases as your sample size increases. For example, to increase a sample from two fifty to one thousand only doubles the precision. You must make a decision about your sample size based on factors such as time available, budget and necessary degree of precision.

By keeping in mind the objectives of the survey, the data should be collected. One should not collect too many data, which are never examined and analyzed. In practical situation, one can chalk out an outline of the table, which would help us in eliminating the collection of irrelevant information.

3. Tools for Collection of Information

5. Tools for collection of information
Questionnaire or Schedule?
After deciding about the type of the data to be collected the next important part of the sample survey is the construction of the questionnaire which has to be filled by the respondents or a schedule which has to be completed by the investigator.

This tool of data collection is quite popular in case of big enquiries. It is adopted by private and public organizations, research workers, individuals and even by the governments. A questionnaire consists of number of questions printed or typed in a definite order on a form or set of forms.

Quite often questionnaire is considered as the heart of a survey operation. Hence, it should be very carefully constructed. If it is not properly set up then the survey is bound to fail. This fact requires us to study the main aspects of questionnaire that is a general form, a question sequence and question formulation and wordings.

This requires familiarity with the subject matter, skill and special techniques. The first rule is to design the questionnaire to fit the medium.
KISS - keep it short and simple. If you present a twenty-page questionnaire most potential respondents will give up in horror before even starting. Ask yourself what you will do with the information from each question. If you cannot give yourself a satisfactory answer, leave it out.

Avoid the temptation to add a few more questions just because you are doing a questionnaire anyway. If necessary, place your questions into three groups: must know, useful to know and nice to know. Discard the last group, unless the previous two groups are very short.

Start with an introduction or welcome message. In the case of mail or Web questionnaires, this message can be in a cover page or on the questionnaire form itself. If you are sending emails that ask people to take a Web page survey, put your main introduction or welcome message in the email.

When practical state - who you are and why you want the information in the survey. A good introduction or welcome message will encourage people to complete your questionnaire.

Allow a “Don't Know” or “Not Applicable” response to all questions, except to those in which you are certain that all respondents will have a clear answer. In most cases, these are wasted answers as far as the researcher is concerned, but are necessary alternatives to avoid frustrated respondents.

Sometimes “Don't Know” or “Not Applicable” will really represent some respondents' most honest answers to some of your questions. Respondents who feel they are being coerced into

giving an answer they do not want to give often do not complete the questionnaire. For example, many people will abandon a questionnaire that asks them to specify their income, without offering a "decline to state" choice.

For the same reason, “Other” or “None” are also included whenever either of these is a logically possible answer. When the answer choices are a list of possible opinions, preferences, or behaviours, you should usually allow these answers. On paper, computer, direct and Internet surveys are the four choices, which should appear as appropriate.

You may want to combine two or more of options into one choice, if you have no interest in distinguishing between them. You will rarely want to include “Don't Know,” “Not Applicable,” “Other” or “None” in a list of choices being read over the telephone or in person, but you should allow the interviewer the ability to accept them when given by respondents. Make sure your questions accept all the possible answers.

A question like "Do you use regular or premium gas in your car?" does not cover all possible answers. The owner may alternate between both types. The question also ignores the possibility of diesel or electric-powered cars.

A better way of asking this question would be "Which type(s) of fuel do you use in your cars?" The responses allowed might be: a) Regular gasoline b) Premier gasoline c) Diesel d) Other e)do not have a car.

Hence, the questions should be clear and short, in such a way that there should not be any scope for guessing for both and also should not hurt the feelings of the respondents.

4. Methods of Collection of Information

5. Methods of collection of information

There are several methods for the collection of primary data particularly in surveys and descriptive researches. Some of them are:

- Observation method
- Interview method
- Through questionnaire
- Through schedule etc.

Most popular methods of collection of information are:
Interview method and Mail questionnaire.

a) Interview method: Investigator goes from house to house and interviews the individual's personally. He fills up the schedule in accordance with the answers provided by the respondents

b) Mail Questionnaire Method: This is a method in which a questionnaire is posted to the individuals who are required to fill up and return after completing in a specified time

Any one of the above two methods can be adopted for collecting the information depending upon cost and time.

Mailed method costs less but leads to considerable non response. This method will work out only among the educated persons. The method of collecting data by mailing the questionnaire to the respondents is most extensively employed in various economic and business surveys.

On the other hand, interview method costs more and there will be an interviewer's error. For the successful implementation of the interview method, interviewers should be carefully selected, trained and briefed. They should be honest, sincere, hardworking, and impartial and must possess the technical competence and the necessary practical experience. In fact, interviewing is an art governed by certain scientific principles.

Sometimes data can't be collected for all the sampled units.
For example, the selected respondent may not be available or they may fail or even refuse to provide information to the investigator when they are contacted. This incompleteness is called non response which will tend to change in results that have to be handled with great care in order to draw valid conclusions.

6. Selection of the proper sampling design

The size of the sample, the procedure for selection of the sample and estimation of population parameters along with the margin of error are some of the important statistical problems that

should receive most of our attention. A number of designs for the selection of the sample are available and one has to select judiciously which will guarantee reliable estimates.

7. Organization of the field work

Personnel involved in the enquiry must be well trained in locating the sampling units, recording the measurements, the methods of collection of the required data before starting the fieldwork.

8. Pre-test the Questionnaire

The last step in questionnaire design is to test a questionnaire with a small number of interviews before conducting your main interviews. Before a survey is committed to the field, the various procedures and designs of the survey should be tested.

Ideally, you should test the survey on the same kind of people that you will include in the main study. If that is not possible, at least have few people, other than the question writer and try the questionnaire. This kind of test run can reveal unanticipated problems with question wording, instructions to skip questions, etc. It can help you see if the interviewees understand your questions and give useful answers.

If you change any questions after a pre-test, you should not combine the results from the pre-test with the results of post-test interviews. The Survey System will invariably provide you with mathematically correct answers to your questions, but choosing sensible questions and administering surveys with sensitivity and common sense will improve the quality of your results dramatically. Pilot surveys are the procedures that are used to test and refine the survey before it is actually fielded.

Small Pre-test (Testing the questionnaire and field methods on a small scale) is very useful. It always helps to decide upon the effective method of asking questions and in improving the questionnaire. It discloses certain problems and troubles that will be very serious on the large scale.

5. Analysis of the Data

9. Summary and Analysis of the data

The analysis may be broadly classified as:

i. Scrutiny and editing of the data
ii. Tabulation of data
iii. Statistical Analysis
iv. Reporting and Conclusion

I. Scrutiny and Editing of the data:

When the investigators are on the field, the supervisory staff should carry out initial quality check. Accordingly, the filled questionnaire or the schedule should be thoroughly scrutinized and examined for the consistency of the data obtained.

This will help in eliminating the erroneous data. Most research studies results in a large volume of raw data, which must be suitably reduced so that the same can be read easily and can be used for further analysis.

II. Tabulation of data:

Before going for the tabulation of data, one has to decide about the methods of tabulation, which are incomplete due to non response. There are two types of tabulation. They are:

i) Manual tabulation
ii) Machine tabulation

The choice depends upon the quantity of the data available.

Classification and tabulation achieves the objective of reduction of data to some extent but we have to go a step further and develop certain indices or measures to summarize the collected classified data. Only after this we can adopt the process of generalization from samples to population.

III. Statistical Analysis (Principal findings)

Different methods of estimation are available for the same data. Appropriate formula must be used to provide the final estimates of the required information.

Most of the surveys have a key requirement, which needs to be a representative of the population of interest. In fact there are two major areas of Statistics they are:

1) Descriptive Statistics
2) Inferential Statistics

Descriptive statistics are concerned with the development of certain indices from the raw data whereas, the inferential statistics concerns with the process of generalization. Inferential statistics are also known as sampling statistics and are mainly concerned with two major types of problems:

i. Estimation of population parameters
ii. The testing of Statistical hypothesis

The important statistical measures that are used to summarize the survey data are:

- Measures of central tendency or averages
- Measures of dispersion
- Measures of asymmetry (skewness)
- Measures of relationship and others

In statistical terms the representativeness can be defined in the following ways:

- Sample means are statistically not different from the population means
- Sample variances are statistically not different from the population variances
- Sample covariance's are statistically not different from the population covariance's

Organize Results—Results are organized into categories in tables, charts, graphs, maps, and/or other written forms making appropriate calculations (e.g. total growth, distances, total number observed).

Analyze Data to Look for Patterns and Trends—Populations are estimated; means, modes, medians, t-values and r-values are calculated; graphs, tables, or maps are analyzed for patterns; data are compared to standards.

IV. Reporting and Conclusion:

Finally, a report consisting detailed statement of the different stages of the survey should be prepared.

10. Information gained for future surveys

A completed survey will always provide us a note of causation and a lesson for designing the future surveys. Information gained in the previous surveys will help in identifying the nature of variability of the measurements together with the cost involved in obtaining the data and serves as a guide for improved future surveys.

Here's a summary of our learning in this session:

- Discussed about sample design
- Principal steps followed in a sample survey
- Tools and techniques used for the collection of information
- Statistical Analysis or principal findings from the data