

Unit 23

Survey Execution and Data Analysis

Contents

- 23.1 Introduction
- 23.2 Problems and Issues in Executing Survey Research
- 23.3 Data Analysis
- 23.4 Ethical Issues in Survey Research
- 23.5 Conclusion

Learning Objectives



It is expected that after reading Unit 23 you would be able to

- ❖ Identify the difficulties one faces in executing a survey
- ❖ Carry out the various steps of data analysis, namely, editing, coding, tabulation and data analysis and finally report writing
- ❖ Recognise the kind of ethical and moral issue the researcher faces.

23.1 Introduction

A number of issues spring up throughout the process of executing survey research. We will examine a few of them in order to enhance the quality and reliability of survey. The researcher may confront/encounter many more situations that may inspire him or her to serious thinking about the implementation of the survey plan. The best way to tackle any problem in the field is to keep oneself open to surprises that may come.

23.2 Problems and Issues in Executing Survey Research

There are a number of problems a researcher faces after the preparation and plans for survey have been charted out. The problems may arise out of the poor training of interviewers or investigators, locating and accessing the respondents and field settings, implementing the data collection techniques, and in terms of the costs and resource management. This section presents some of the issues that may confront the researcher in the field.

i) Locating and accessing the field settings

Once the training of the interviewers is over, the actual process of reaching out to respondents and their field settings begins. The first task of the researcher at this stage is to identify the target population and write to the respective organisations and institutions for permission to conduct survey. It is at this stage that one must keep in mind that there are two different kinds of field settings, namely, formal (official) and informal

(unofficial settings). Thus, each of the settings would have some people who may be called the 'gatekeepers', the people with the authority to permit entry into the field settings.

a) Gatekeepers in a formal organisation

Official gatekeepers are the ones who give permission to conduct survey in formal organisations such as schools, government departments, NGOs, etc. (Headmaster in a School or a District Educational Officer at the district level or the Director of an NGO). If we enter an organisation without the permission of the gatekeepers, our survey may face threats of legal action for not taking permission as it is the right of the institution to participate in the study or not. The researcher has to ensure the approval and consent of the gatekeepers before entering formal institutions, such as schools, government departments, NGOs, hospitals, etc.

b) Gatekeepers in an informal setting

It is just not merely in the formal settings, in informal settings also (such as family, community, etc.) gatekeepers are involved. One requires permission to conduct survey research. If the researcher wishes to speak to a child in the home environment, he must speak to the parents first. Usually the parents advise their children to speak to the surveyors. If the researcher intends to speak to women of a house, he is expected to take the permission of the male head of the family. When the researcher wants to speak to a community member, he has to speak to the head of the community (village head, panchayat chairman, caste leaders, etc.). Without the permission of the gatekeepers in informal settings, the members of a family or a community may not cooperate or may even turn down the request for interview. Therefore, the interviewer must be conscious of the importance of gatekeepers in informal settings as well.

c) How do we approach the gatekeepers for permission?

The first thing a researcher must do after identifying the respondents and their settings is that they should inform the gatekeepers about the objectives and purposes behind the survey and apprise them of its overall goal. For instance, the main objective of a survey being undertaken is to assess the impact of the Integrated Education for the Disabled Children. The respondents may be informed that the survey results may be used to improve the quality of implementation and delivery of the scheme and that their cooperation is highly appreciated and valuable. Further, the researcher has to clarify the doubts and apprehensions of both the gatekeepers and respondents. For this, the researcher must be well-versed with the specific objectives of the survey. For instance, the broad objectives of the survey mentioned above are

- ❖ To study how far the integrated schools have been successful in integrating the eligible children with special needs.
- ❖ To examine the retention strategies adopted by the schools.

- ❖ To study whether the facilities for assessment and appropriate placement of the children with disabilities are available in the integrated schools.

There can be many specific objectives of the current survey. So the interviewer should try to understand these objectives and be clear about each of them. This will help him in the later stages of survey execution. It will help to correctly transact and impart the same kind of training to his assistants. The interviewer must also understand the survey's conceptual, cultural, psychological, economic, and political context.

Some of the specific questions the researcher/interviewer must be clear about in the case of our example are

- ❖ Who are the children with special needs?
- ❖ What are their special needs?
- ❖ What is the need for a special programme for children with special needs?
- ❖ What is the cultural context of the families, communities and the schools that are going to be a part of the study?
- ❖ What will be the psychological condition of the children with disabilities or special needs and their parents?
- ❖ What is the economics of running a scheme such as this (IEDC)? What are the possible uses of equipment and their utility and how they are used? Whether the investment on certain items is justified in terms of their utility.
- ❖ What is the political context of the schooling of children with special needs? The scheme might be receiving the major focus or less focus in terms of the attention of the implementing agencies, etc. Why has it become so important in recent times? What are the pressure groups asking for the betterment of the scheme? Who are the stakeholders in the programme and what are their positions vis-à-vis the objectives of the scheme?

Further, the researcher must assure the gatekeepers of anonymity and confidentiality of the information provided by them and their organisation. In case one wants to be explicit about the institution / organisation, he should request permission for doing so. Once he is in the field, after gaining entry or access to the setting, a researcher must gain access to the individual respondents. Some important tips to access the respondents are

- ❖ Accessing the respondents and building rapport with them:
Remember the interview situation is one of social interaction. In a social interaction, building rapport is essential if you want your relationship to sustain. Rapport is nothing but building relationships.
- ❖ Approach the respondents for permission to include them in the

study. The respondent must be willing to participate in the survey. Everybody in the sample has a right to decide whether he or she would like to be part of the survey. If one is forced to participate in the survey, then the reliability of the information is at stake. There is also the possibility of the respondent providing misleading, incomplete, or incorrect answers.

- ❖ Do not think that it is the privilege of the respondents to give the researchers the interviews and they must cooperate with him. So, the researcher must understand what matters to the respondents before including them in the study (Often surveys are criticized for being driven by the interests of the researcher).
- ❖ Recognise the differences between the respondents: If all the respondents are alike in thought and action, then there is no need for conducting a survey. We can simply take a case and generalize from that case. Variations make the findings of the survey worthwhile and interesting.
- ❖ Avoid unjustified presuppositions and false assumptions about the respondents: Often, we are carried away by our presuppositions and stereotypes about the other individuals. For instance, one may view children with disabilities or special needs differently. This will lead to a bias in the way the questions are asked and even the respondents are accessed. Simply put, the interviewer should not be judgmental in his interaction with the respondents as also with the topic of the survey, the institutions they approach for information, or even the role of the parents, etc.
- ❖ Be a good listener, rather than a good speaker.

Reflection and Action 23.1

Suppose you need to carry out a survey of waste disposal methods of your area, what would be your strategies for reaching out to the respondents? Write a short note on the qualities you would like to see in the interviewers appointed for the survey.

ii) Implementation of data gathering technique

There are a number of problems which spring up after entering the field while executing the main survey. The problems mainly concern the implementation of the data gathering techniques such as self-administered questionnaires and face-to face interviews.

a) Self-administered questionnaires

A major problem with the self-administered mailed questionnaires is how to get them returned. We already know that the respondents face considerable temptation not to fill up the questionnaires mailed to them. The suggestion offered in such cases is that a 50 percent return rate be used as the minimum level of acceptability for survey research based on the self-administered questionnaires. This level can be achieved in some

situations such as those in which a researcher distributes questionnaires personally in a classroom, office premises, etc., but rarely when data collection is based on a mail survey. However, some steps may be suggested to maximize the return rate.

- ❖ An introductory letter should accompany the mailed questionnaire. This letter should explain the purpose of study and who is sponsoring the project. A good introductory letter can be an incentive for the respondent to fill out the questionnaire.
- ❖ Besides an introductory letter, the inclusion of a stamped, self-addressed envelope for the respondent to return the questionnaire is a necessity if a good return rate is to be achieved.
- ❖ In some cases, special incentives such as a pen or a small amount of money are also sent to the potential respondent in order to encourage cooperation. However, this step may raise some ethical questions, besides being costly.
- ❖ Finally, what is crucial to a good return rate is an intensive follow-up campaign which attempts to encourage those who have not returned their questionnaires by a certain date.

Sometimes, even if all these steps for increasing the return rate of the self-administered questionnaires are employed, it is unlikely that more than seventy percent of the questionnaires will be returned. The question then remains: How does the thirty percent who do not respond, affect the research outcomes? Can we say that the representativeness of the sample is not jeopardised? The answer is both Yes and No.

For instance, sometimes the poor return rate might affect the research findings as a result of the lack of representativeness. At other times, it may not at all affect the findings of a survey. One way of dealing with this question is to compare the information about those who do not respond with information derived from some other source about the sample. For example, assume that a self-administered questionnaire is mailed to a simple random sample of the students of a college and that sixty percent of the sample return their questionnaires. Information from the survey about the respondent's age, sex, and subject of study can be compared to the information derived from the materials available on the total student body. If there are no startling differences between the survey data and the data from other sources, it could be concluded that the forty percent of the sample which did not return their questionnaire is similar to the sixty percent of those who did return them. In this case the issue of representativeness is favourably resolved. However, it is important to mention that this is not a perfect solution and there can be other ways of resolving it.

b) Face-to-face interviews

Even though the response rate is considerably higher for interviews, there are similar problems involved in face-to-face interviews. Some

strategies for improving the response rate in interviews are the following:

- ❖ The refusal of cooperation by potential respondents can be lessened by first training the interviewers in ways to establish rapport with the respondents.
- ❖ The temptation of the interviewers to avoid certain respondents should be diminished by emphasising the validation procedure and requiring the interviewers to account for each non-completion. For example, sometimes a researcher doing door-to-door survey may avoid a house that displays a board, "Beware of the Dog".
- ❖ Respondents are often not home when the researcher attempts to contact them. It is therefore important to make visits at different times of the day and on different days of the week.

If these strategies are used effectively and the response rate is average or better, there is still the question of how representative the respondents are of the total group one has tried to study. The suggestion given in the case of self-administered questionnaire for comparing the survey information on the respondents with other information about the total group is applied in the case of face-to face-interviews as well.

Further, a major issue crucial for the success of the implementation of the interview process is the selection and training of the interviewing staff. The following aspects should be kept in mind while training the interviewers:

- ❖ Train interviewers in how to establish rapport with respondents in order to gain their cooperation. However, rapport does not mean to encourage the pleasing answers.
- ❖ The interviewers must be told to remember that the interview situation is one of social interaction. Differences in characteristics such as caste, race, sex, social class, religion) of both the respondent and the interviewer have an impact on the answers elicited.
- ❖ The interviewers should become non-judgmental towards the respondents. This quality should make the respondents feel comfortable in expressing their true feelings.
- ❖ It makes good sense to hire and train interviewers who are not radically different from the people whom they will be interviewing.
- ❖ A manual may be prepared for the interviewers which deals with the general interviewing techniques and the unique requirements of the particular study in question.

After the actual interviewing process begins, interviewers should be asked to return their first few completed interviews to the supervisory personnel for critical evaluation. At this point, some validation interviews should be conducted by telephone or in person to ensure that the interviewers

are not irritating any respondents and that they are really conducting the interviews according as the instructions. Such an assessment is helpful for the purpose of evaluating the questionnaire under actual field conditions. Independent validation of the interviews should continue throughout the time-period of the interviews. See Box 23.1 for guidelines.

Box 23.1 Some guidelines for effective implementation of the data gathering techniques

The researcher has to keep in mind the following guidelines to avoid some problems in the collection of information from the respondents.

- ❖ If your questionnaire is in English and your respondents are not comfortable with English, then try to translate the questions in the regional languages while asking the respondents.
- ❖ Make the questions meaningful, clear, unambiguous, sensitive, and revealing. The use of language also depends on factors such as age, region, social class, etc.
- ❖ Understand the social standing of particular usages in language. For instance, the 'mid-day meal' is 'lunch', not 'dinner', the room where the family gathers is the 'sitting room', not the 'living room', a 'magazine' is never called a 'book.'
- ❖ Sometimes people may think you are being impolite or vulgar or rude or incorrect based on the kind of language you use while interviewing them. This may spoil the entire process of interview itself. So use proper language.
- ❖ Tactics for dealing with ambiguous or unclear terms:
 - Avoidance. Use alternatives
 - Another possibility is to gloss the term, i.e. to give a brief explanation to the respondent of what we mean by the word. But one problem with glossing is that we almost impose the meaning or even the answer on the respondent.
 - Clarification. This is a form of glossing, but we explicitly clarify the potential ambiguity
 - Giving Examples. But, it carries the danger of suggesting the answers or may even distract the respondents from thinking about other answers.

23.3 Data Analysis

Once the data have been collected, the task is to reduce the mass of data obtained to a form suitable for analysis. 'Data reduction[Ⓢ]', as the process is called, generally consists of editing, coding and processing through the computer in the case of large surveys and by hand in the case of small surveys.

a) Editing

Editing the filled in questionnaires is the first step in the preparation of data for processing. It is important to make sure that the questionnaire has been filled up properly. This is a painstaking work but it can be very fruitful. In sum, a good editing job can add substantially to the quality of the collected data.

Editing interview schedules or self-administered questionnaires is intended

to identify and eliminate errors made by the interviewers and the respondents. Moser and Kalton (1973), speak of three tasks in editing:

- ❖ **Completeness:** A check is made that there is an answer to every question. In most surveys, interviewers are required to record an answer to every question. Missing answers can sometimes be cross-checked from other sections of the survey. At worst, respondents can be contacted again to supply the missing information.
- ❖ **Accuracy:** As far as possible, a check is made that all questions are answered accurately. Inaccuracies arise out of carelessness on the part of either the respondents or the interviewers. Sometimes a deliberate attempt is made to mislead. A tick in the wrong box, a ring around the wrong code, an error in simple arithmetic can reduce the validity of the data unless they are picked up in the editing process.
- ❖ **Uniformity:** A check is made that interviewers have interpreted instructions and questions uniformly. Sometimes the failure to give explicit instructions over the interpretation of respondents' replies leads to interviewers recording the same answer in a variety of answer codes instead of one. A check on uniformity can help eradicate this source of error.

b) Coding

The quantitative analysis of survey data requires that answers are converted into numbers. Many variables also require that answers be classified into categories. This process of converting answers to numbers and classifying answers is called Coding[®]. Thus, the primary task of data reduction is coding, which is primarily assigning a code number to each answer to a question in the questionnaire / interview schedule. There are six main steps in coding and classifying questionnaire data. They are:

- ❖ Classifying responses
- ❖ Allocating codes to each variable
- ❖ Allocating column numbers to each variable
- ❖ Producing a Codebook
- ❖ Checking from coding errors
- ❖ Entering data

Let us discuss in brief each step in coding and classifying data.

i) **Classifying responses:** Coding is more than allocation of numbers to the responses. It also involves the creation of a classification system that imposes a particular order on the data. This in turn affects the way data are analysed. These classification systems are not objective systems but are created by people and reflect the historical and cultural ways in which we make sense of the world around us. David de Vaus (2002) argues that as creators and consumers of research we need to be aware

that classification systems shape what we find.

ii) **Allocating codes to each variable:** Once classification schemes have been developed, our task is to allocate codes to each of the categories in the classification. These classification schemes can be developed either before a questionnaire / interview schedule is administered or after. Much of the work in the classifying responses is undertaken at the questionnaire construction stage where a set of fixed responses is provided to the respondents. Codes are allocated to these responses and these codes are normally visible in printed questionnaires to assist with data entry at a later stage.

Open-ended questions are coded after the data have been collected. Post-coding is done either by using systematic, pre-existing standard coding schemes or developing a coding scheme based on the responses provided by the respondents. For example, standard questions such as occupation, religion, caste, type of family, etc., can have standardized coding schemes. The standardized coding schemes are systematic and are developed by people with considerable expertise after considerable consultation. Further, they are publicly available and reduce code error.

Each variable has at least two categories and any person must belong to one and only one category. The essence of coding is to give a distinctive code to each category of a variable.

iii) **Allocating column numbers to each variable:** To enter codes into a computer they must be put on a 'record'. In the early days of computing a record was a computer card that could contain up to 80 digits. If more than 80 digits were required for all the respondent's answers a second card or record would be used for that case. These days with electronic recording, data can be much longer than eighty digits. The records for each case are then placed in a data file in which the first record represents the first respondent, the next record represents the second respondent and so forth.

iv) **Producing a codebook:** After deciding how to code each response to each variable, it is important to make a systematic record of all the decisions made. This record is called a Codebook and the following information is normally included.

- ❖ The exact wording of the question
- ❖ A name by which a variable is referred to in the programme. In the Codebook list a name is given to each variable.
- ❖ The type of data used for that variable.
- ❖ The first and last column numbers in which the variable is located.
- ❖ The valid codes for each question.
- ❖ The missing data codes for each question.
- ❖ Any special coding instructions used for coding particular questions.

- ❖ **Checking for coding errors:** Coding errors can create serious problems during data analysis. The most serious errors figure when data are entered in the wrong columns. But miscodes are a more common problem. These can occur during the data collection phase, during the manual coding of the answers or during the data entry phase. It is probably impossible to eliminate all coding errors but the problem can be reduced by locating and correcting as many errors as possible.
- ❖ **Entering data:** In the past, data were always coded manually and entered into the computer by keypunch operators. However, advances in computer technology have radically changed the whole process of data entry.

c) Tabulation and data analysis

By a simple command the computer gives all the tabulations that are anticipated for data analysis. It is the simplified version of the entire data collected, which is given in numerical values of the distribution of responses of the respondents. A researcher labours hard through several stages of survey research to arrive at this stage. The stage of tabulation sums up the outcome of the survey research in a skeleton form. However, it is important that the researcher assumes some knowledge of quantitative data analysis procedures to make sense of the skeleton.

Aldridge and Levine argue that there are three different aspects of a survey analysis. According to them these three dimensions are conceived to be potentially present in the analysis of any survey although they may not be fully or equally exploited. The three dimensions are: descriptive, analytical and contextual. The descriptive dimension tends to dominate in primarily descriptive surveys. The analytic and contextual aspects are more pronounced in analytic surveys, but the art of analysis is to promote the development of all three so that the research potential of a survey is fully realized.

It may also be mentioned here that the data analysis is not just a technical matter. Social scientists have ethical responsibilities to analyse data properly and report it fairly. See Box 23.2 for what you need to be careful about before analysing your data.

Box 23.2 Before Analysis of the Data, Be clear about Four Broad Considerations

- ❖ What is the level of measurement of each variable being used in any particular piece of analysis?
- ❖ How many variables will any particular piece of analysis require?
- ❖ What type of analysis is required? Is descriptive analysis required or analytic or contextual analyses?
- ❖ Have the ethical principles of full, fair, appropriate and challenging analysis been applied to the selection of data to be analysed and reported?

d) Report writing

Once the data are available the survey researchers typically write and disseminate reports or present them orally. In any case, the presentation of the results of the survey should be as clear and complete as possible. The following are some questions that are useful for evaluating the quality of a survey research report of findings.

- ❖ Could the reader replicate the report on the basis of the information provided? Are all the phases of survey research process explained in adequate detail?
- ❖ Are all the potentially relevant data provided or only highly selected excerpts included? Most studies call for much more in the way of preliminary statistical analysis than it would be appropriate to include in the final report. Thus some selectivity is always necessary. But we can question the nature of the selectivity when data for what seem to be superior indicators of the phenomena of interest are not presented and no rationale for their exclusion is provided.
- ❖ Are all the data provided supportive of the researcher's hypothesis? When there is no data to contradict the researcher's original hypothesis, we might question the selection procedure used in deciding which results to present. Also related is the tone of the report. Is it one of justifying the hypothesis or of critically evaluating it?
- ❖ Are the study's conclusions warranted on the basis of the data that have been presented? It is not uncommon for investigators to inflate the substantive significance of their data to the point that the reader of the text is led to the conclusion that the relationship is very strong when the actual statistical analysis yields a weak or moderate relationship. It is not wise to skip the tables in a report and rely on the author's conclusions about the content of these tables as presented in the text of the report.
- ❖ Have the authors demonstrated a desire to put their hypotheses in jeopardy and to really dig up any contradictory evidence that may exist? Or in contrast, do the authors seem to be interested only in presenting results that support their hypothesis?
- ❖ Is there any discussion of the author's values and the ways in which they may have had an influence on the outcome of the study?

As more and more research is conducted, an ever-increasing amount of data becomes available to researchers. It is often possible to obtain the data desired simply by reanalysing data reproduced by someone else. This is called secondary analysis of data and it can be used to substantiate or refute the earlier researches on any topic of investigation during the process of report writing.

23.4 Ethical Issues in Survey Research

Like all social research, people can conduct surveys in ethical or unethical ways. A major ethical issue in survey research is the invasion of privacy. Survey researchers can intrude into a respondent's privacy by asking about intimate actions and personal beliefs. People have a right to privacy. Respondents decide when and to whom to reveal personal information. They are likely to give such information when it is asked for in a comfortable context with mutual trust, when they believe answers will remain confidential. Researchers should treat all the respondents with dignity and reduce their anxiety or discomfort. They are also responsible for protecting the confidentiality of data.

The second ethical issue that confronts the survey researcher is the right of respondents to participate voluntarily. Respondents agree to answer questions and can refuse at any time. They give 'informed consent' to participate in research. Researchers depend on the respondents' voluntary cooperation. Therefore, researchers need to ask well-developed questions in a sensitive way, treat respondents with respect, and be very sensitive to confidentiality.

Thirdly, the tendency to mislead the people at large by the survey researchers is also an important ethical consideration. Sometimes, people may also misuse the survey results or use poorly designed or purposely rigged surveys. People may demand answers from surveys that surveys cannot provide or may not understand a survey's limitations. Those who design and prepare surveys may lack sufficient training to conduct a legitimate survey. Sometimes the researchers use the survey format in an attempt to persuade some interviewer to conduct survey, with little or no real interest in learning information from a respondent.

Reflection and Action 23.2

Suppose you have to guide a team carrying out a survey of methods of waste disposal in your area. Explain how you will handle the following questions.

Questions

- ❖ What would be your strategy to avoid the policy decisions based on careless or poorly designed surveys may result in waste and human hardship?
- ❖ How would you make the survey execution so that legitimate researchers conduct methodologically rigorous survey research?
- ❖ How would you make researchers aware of and report the limitations of the survey research?
- ❖ How can researchers combat unscrupulous politicians, business people, and others who rig surveys to produce deceptive results?

Write on a sheet of paper your answers to the above questions and add at least four more ethical issues that await a researcher while conducting a survey. Some of them may deal with the sponsorship and funding of the project, researcher's personal preferences and biases, the respondent's deception, etc. Finally answer the question: Who is responsible for dealing with the problem areas and avoiding the ethical issues in order to make the research endeavour successful and objective?

23.5 Conclusion

In this unit, we have attempted to learn the issues and problems that researchers confront while executing the survey. Primarily, it discusses the issue of locating and accessing the field settings and respondents and the problems faced during the administration of data gathering techniques. The unit also discusses the phase of analysis after the completion of data collection and the execution of the main survey. Finally, the ethical issues in the conduct of survey have been discussed.

It may be summed up that survey is a process in which the researcher translates a research question into a questionnaire, which then is used to collect data from respondents. The data thus collected is processed through a rigorous process of editing, coding, and computerisation and tabulation that emerge in the end. This data in turn attempts to analyse the research question formulated at the beginning of the survey research. The entire process requires the utmost attention of the researcher in order to avoid and minimize errors, which may distort the very purpose of research, i.e., search for truth.

Further Reading

Cohen, Louis and Manion, Lawrence. 1994. *Research Methods in Education*. Routledge: London and New York

Neuman, W. Lawrence. 1997. *Social Research Methods*. Allyn and Bacon: Boston

Glossary

(Explanations of glossary words have been prepared with the help of information available on the Internet and in other sources.)

Chi-square: Chi square is a non-parametric test of statistical significance for bivariate tabular analysis (also known as cross breaks). Any appropriately performed test of statistical significance lets you know the degree of confidence you can have in accepting or rejecting a hypothesis. Typically, the hypothesis tested with chi square is whether or not two different samples (of people, texts, whatever) are different enough in some characteristic or aspect of their behaviour that we can generalize from our samples that the populations from which our samples are drawn are also different in the behaviour or characteristic. A non-parametric test, like chi-square, is a rough estimate of confidence; it accepts weaker, less accurate data as input than parametric tests (like t-tests and analysis of variance, for example) and therefore has less status in the pantheon of statistical tests. Nonetheless, its limitations are also its strengths; because Chi Square is more 'forgiving' in the data it will accept, it can be used in a wide variety of research contexts.

Coding: A systematically arranged and comprehensive collection of laws or a systematic collection of regulations and rules of procedure or conduct or a system of signals used to represent letters or numbers in transmitting messages or a system of symbols, letters, or words given certain arbitrary meanings, used for transmitting messages requiring secrecy or brevity.

Cohort: Cohort derives from Latin *cohors*, "an enclosure, a yard," In Statistics 'cohort' means a sample meant to be representative of a whole population. In informal usage it refers to a variety or diversity.

Concept of probability: Inferential statistics is built on the foundation of probability theory, and has been remarkably successful in guiding opinion about the conclusions to be drawn from data. Yet (paradoxically) the very idea of probability has been plagued by controversy from the beginning of the subject to the present day. One conception of probability is drawn from the idea of symmetrical outcomes. For example, the two possible outcomes of tossing a fair coin seem not to be distinguishable in any way that affects which side will land up or down.

Correlation coefficient: The Pearson Product-Moment Correlation Coefficient (r), or correlation coefficient for short is a measure of the degree of linear relationship between two variables, usually labeled X and Y . While in regression the emphasis is on predicting one variable from the other, in correlation the emphasis is on the degree to which a linear model may describe the relationship between two variables. In regression the interest is directional, one variable is predicted and the other is the predictor; in correlation the interest is non-directional, the relationship is the critical aspect. The computation of the correlation coefficient is most easily accomplished with the aid of a statistical calculator. The value of r was found on a statistical calculator during the estimation of regression parameters in the last chapter. The correlation coefficient may take on any value between plus and minus one.

Cross-sectional: Adjective of cross-section that relates to a section formed by a plane cutting through an object, usually at right angles to an axis or a piece so cut or a graphic representation of such a piece. In physics, it refers to a measure of the probability that an encounter between particles will result in the occurrence of a particular atomic or nuclear reaction. In statistics it refers to a sample meant to be representative of a whole population.

Diachronic and synchronic: Diachronic is a convenient way of referring to something that changes over time. Synchronic refers to 'similar' instances existing at the same time.

Dichotomous: Dividing into two parts; relating to, involving, or proceeding from a division or the process of dividing into two, especially mutually exclusive or contradictory groups.

Eclecticism: Eclecticism is an approach to thought that does not hold rigidly to a single paradigm or set of assumptions or conclusions, but instead draws upon multiple theories to gain complementary insights into phenomena, or applies only certain theories in particular cases. This is sometimes inelegant, and eclectics are sometimes criticised for lack of consistency in their thinking, but it is common in many fields of study. For example, most psychologists accept parts of behaviorism, but do not attempt to use the theory to explain all aspects of human behavior. Similarly, a physicist may use Newton's laws for predicting the motion of baseballs, but will switch to the relativity for predicting motion of galaxies or to quantum mechanics for the one of subatomic particles. A statistician may use frequentist techniques on one occasion and Bayesian ones on another. An example of eclecticism in economics is John Dunning's eclectic theory of international production. Eclecticism in psychology is also supported by many in that in reality many factors influence behaviour and psyche, therefore it is inevitable to consider all perspectives in identifying, changing, explaining, and determining behaviour.

Eclecticism was first articulated by a group of ancient philosophers who tried to select from the existing philosophical beliefs those doctrines that seemed most reasonable to them. Out of this collected material they constructed their new system of philosophy. The term comes from the Greek *eklektikos*: choosing the best. Well known Eclectics in Greek philosophy were the Stoics Panaetius and Posidonius, and the New Academics Carneades and Philo of Larissa. Among the Romans, Cicero was thoroughly eclectic, as he united the Peripatetic, Stoic, and New Academic doctrines. Further eclectics were Varro and Seneca.

Ex-post-facto: Ex post facto is Latin for "from a thing done afterward." Ex post facto is most typically used to refer to a law that applies retroactively, thereby criminalizing conduct that was legal when originally performed.

Gatekeeper: One that is in charge of passage through a gate. The term refers to one who monitors or oversees the actions of others.

Homogeneity: In statistics, homogeneity is the concept related to reliability. While the internal consistency reliability (statistics) indicates the degree a data set approximates an ideal equivalence scale, the homogeneity indicates the degree a data set approximates an ideal.

Hypothesis testing: Setting up and testing hypotheses is an essential part of statistical inference. In order to formulate such a test, usually some theory has been put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proved, for example, claiming that a new drug is better than the current drug for treatment of the same symptoms. In each problem considered, the question of interest is simplified into two competing claims / hypotheses between which we have a choice; the null hypothesis, denoted H_0 , against the alternative hypothesis, denoted H_1 . These two competing claims / hypotheses are not however treated on an equal basis: special consideration is given to the null hypothesis.

Informal Consent: Casual or in an informal manner or an informal agreement to or acceptance of opinion or a course of action.

Information society: An information society is one in which the creation, distribution and manipulation of information is becoming a significant economic and cultural activity. The knowledge economy is its economic counterpart whereby wealth is created through the economic exploitation of knowledge. The information society is a new kind of society. Specific to this kind of society is the central position information technology has for production and economy. Information society is seen as successor to industrial society. Closely related concepts are post-industrial society (cf Daniel Bell), post-fordism, post-modern society, knowledge society, Telematic Society, Information Revolution, and informational society (cf Manuel Castells).

One of the first people to develop the concept of the information society was the economist Fritz Machlup. In 1933 Machlup began studying the effect of patents on research. His work culminated in the breakthrough study "The production and distribution of knowledge in the United States" in 1962. This book was widely regarded and was eventually translated into Russian and Japanese.

Jargon: The specialised or technical language of a trade, profession, or similar group. Also refers to speech or writing having unusual or pretentious vocabulary, convoluted phrasing, and vague meaning.

Logitudinal: The word comes from the Latin *longitudo* meaning length. Hence, longitudinal means along the length, running lengthwise, or (by extension) over the course of time.

Mean: The arithmetic mean is what is commonly called the average: When the word "mean" is used without a modifier, it can be assumed that it refers to the arithmetic mean. The mean is the sum of all the scores divided by the number of scores. The formula in summation notation is: $\bar{x} = \sum X/N$ where \bar{x} is the population mean and N is the number of scores. If the scores are from a sample, then the symbol M refers to the mean and N refers to the sample size. The formula for M is the same as the formula for \bar{x} . The mean is a good measure of central tendency for roughly symmetric distributions but can be misleading in skewed distributions since it can be greatly influenced by extreme scores. Therefore, other statistics such as the median may be more informative for distributions such as reaction time or family income that are frequently very skewed.

Measure of variability: The terms variability, spread, and dispersion are synonyms, and refer to how spread out a distribution is. There are four frequently used measures of variability, the range, interquartile range, variance, and standard deviation.

Median: Being in the middle or in intermediate position; relating to or constituting a statistical median

Mode: The most frequent value of a set of data; a value of a random variable for which a function of probabilities defined on it achieves a relative maximum

Parametric: A parameter is a measurement or value on which something else depends. For example, a parametric equaliser is a tone control circuit that allows the frequency of maximum cut or boosts to be set by one control, and the size of the cut or boost by another. These settings, the frequency and level of the peak or trough, are two of the parameters of a frequency response curve, and in a two-control equaliser they completely describe the curve. More elaborate parametric equalisers may allow other parameters to be varied, such as skew. These parameters each describe some aspect of the response curve seen as a whole, over all frequencies. By way of contrast, a graphic equaliser provides individual level controls for various frequency bands, each of which acts only on that particular frequency band.

Pseudonym: A pseudonym (Greek: false name) is a fictitious name used by an individual as an alternative to their legal name (whereas an allonym is the name of another actual person assumed by one person in authorship of a work of art; e.g., when ghostwriting a book or play, or in parody, or when using a front such as by screenwriters blacklisted in Hollywood in the 1950s, 1960s, and 1970s). In some cases, the pseudonym has become the legal name of the person using it. Occasionally, a pseudonym is employed to avoid overexposure. Prolific authors for pulp magazines often had two and sometimes three short stories appearing in one issue of a magazine; the editor would create several fictitious author names so that readers would not realise this.

Positivism: In sociology, anthropology, and other social sciences, the term was closely connected to sociological naturalism and can be traced back to the philosophical thinking of Auguste Comte in the 19th century. Describing positivism in the 1966 Henry Myres Lecture, the structural anthropologist Edmund Leach said, "Positivism is the view that serious scientific inquiry should not search for ultimate causes deriving from some

outside source but must confine itself to the study of relations existing between facts which are directly accessible to observation." In some quarters of contemporary sociology, positivism has been replaced by antipositivism.

Psephologists: Those who pursue the scientific study of elections

Qualifiers: Qualifiers is a tool for anyone, professional or student, who must assemble, analyse, evaluate, and organise research data. With Qualifiers you can collect all your data—text, audio, video—in one place and arrange it in categories to suit your purpose. Possible qualitative data include conversations, presentations, recitals, field data, and other situations. You can record audio and play back audio and video with tools such as "transcriber play" and "mark back." Perform post-analyses—"intersection," "union," "not in." Print reports. Project analytic presentations of data and trends from your computer.

Range: The range is the simplest measure of variability to calculate, and one you have probably encountered many times in your life. The range is simply the highest score minus the lowest score.

Rapport: The term refers to the relationship, especially one of mutual trust or emotional affinity. It is a French word, from Old French, from *raporter*, to bring back: *re-*, *rè-* + *aporter*, to bring (from Latin *apportare* : *ad-*, *ad-* + *portare*, to carry).

Regression: A functional relationship between two or more correlated variables that is often empirically determined from data and is used especially to predict values of one variable when given values of the others. A classic statistical problem is to try to determine the relationship between two random variables X and Y. For example, we might consider height and weight of a sample of adults. Linear regression attempts to explain this relationship with a straight line fit to the data.

Rigged: To manipulate dishonestly for personal gain

Sample variance: The statistic s^2 is a measure on a random sample that is used to estimate the variance of the population from which the sample is drawn. Numerically, it is the sum of the squared deviations around the mean of a random sample divided by the sample size minus one. Regardless of the size of the population, and regardless of the size of the random sample, it can be algebraically shown that if we repeatedly took random samples of the same size from the same population and calculated the variance estimate on each sample, these values would cluster around the exact value of the population variance. In short, the statistic s^2 is an unbiased estimate of the variance of the population from which a sample is drawn.

Sampling method: It is incumbent on the researcher to clearly define the target population. There are no strict rules to follow, and the researcher must rely on logic and judgment. The population is defined in keeping with the objectives of the study. Sometimes, the entire population will be sufficiently small, and the researcher can include the entire population in the study. This type of research is called a census study because data is gathered on every member of the population. Usually, the population is too large for the researcher to attempt to survey all of its members. A small, but carefully chosen sample can be used to represent the population. The sample reflects the characteristics of the population from which it is drawn.

Standard deviation: The standard deviation is kind of the "mean of the mean," and often can help you find the story behind the data. To understand this concept, it can help to learn about what statisticians call normal distribution of data. A normal distribution of data means that most of the examples in a set of data are close to the "average," while relatively few examples tend to one extreme or the other.

Tabulation: An orderly columnar display of data, like a chart, table, refers to information set out in tabular form, also refers to the act of putting into tabular form.

Triangulation: In trigonometry and elementary geometry, triangulation is the process of finding a distance to a point by calculating the length of one side of a triangle, given measurements of angles and **sides** of the triangle formed by that point and two other reference points. Triangulation is used for many purposes, including surveying, navigation, astrometry, binocular vision and gun direction of weapons. Many of these surveying problems involve the solution of large meshes of triangles, with hundreds or even thousands of observations. Complex triangulation problems involving real-world observations **with** errors require the solution of large systems of simultaneous equations to generate solutions. Famous uses of triangulation have included the retriangulation of Great Britain.

Validity: A deductive argument is sound if and only **if** it is both valid, and all of its premises are actually true. Otherwise, a deductive argument **is** unsound. Deductive argument is said to be valid if and only if it takes a form that makes it impossible for the premises to be true and the conclusion nevertheless to be false. Otherwise, a deductive argument is **said** to be invalid. According to the definition of a deductive argument, the author of a deductive argument always intends that the premises provide the sort of justification for the conclusion whereby if the premises are true, the conclusion is guaranteed to be true as well. Loosely speaking, if the author's process of reasoning is a good one, if the premises actually do provide this sort of justification for the conclusion, then the argument is valid.

Variables: Variables are used in open sentences. For instance, in the formula: $x + 1 = 5$, x is a variable which represents an "unknown" number. In mathematics, variables are usually represented by letters of the Roman alphabet, but are also represented by letters of other alphabets; as well as various other symbols. In computer programming, variables are usually represented by either single **letters** or alphanumeric strings. Why are variables useful? Variables are useful in mathematics and computer programming because they allow instructions to be specified in a general way. If one were forced to use actual values, then the instructions would only apply in a more narrow, and specific set of **situations**.

References

(Please note that the list of references includes all the sources that the authors of the units have cited in the text. It includes also the books mentioned as Further Reading at the end of each unit of Book 2.)

Aldridge, Alan. and Ken, Levine 2001. *Surveying the Social World - Principles and Practice in Survey Research*. Open University Press: Buckingham

Allen G. and G. Skinner 1991. *Handbook for Research Students in Social Sciences*. Falmer Press: London

Babbie, E. 1989. *The Practice of Social Research*. Wadsworth Publishing Company: Belmont, California

Bernard, H. Russel, 1994. *Research Methods in Anthropology: Qualitative and Quantitative Approaches*. Sage: New York

Black, Thomas R. 1999. *Doing Quantitative Research in the Social Science. An Integrated Approach to Research Design, Measurement and Statistics*. Sage Publication: New York

Boas, Franz 1940. The Limitations of the Comparative Method in Anthropology, IN *Race, Language and Culture*. Macmillan: New York

Brewer, John B. 2000. *Ethnography*. Open University Press: Buckingham

Bryman, Alann 1988: *Quantity and Quality in Social Research*. Unwin Hyman: London

Burgess, R.G. (ed.) 1982, *Field Research: A Sourcebook and Field Manual*. (Contemporary Social Research 4), George Allen and Unwin, London

Burns, Robert B. 2000. *Introduction to Research Methods*. Sage Publications: London

Chandler, Dawn and Bill Torbert 2003. *Transforming Inquiry and Action: Interviewing 27 Flavors of Action Research*. *Action Research* 1 (October):

Cohen, Louis and Lawrence, Manion 1994. *Research Methods in Education*. Routledge: London and New York

Cohen, Louis and Michael Holliday 1982. *Statistics for Social Research*. Harper and Row: London

Cubitt, T. ,1973. Network density among urban families. In *Network Analysis: Studies in Human Interaction* (eds J. Boissevain and J.C. Mitchell). The Hague: Mouton

de Vaus, David. 2002. *Surveys in Social Research*. Routledge: London

Denzin, N.K. (ed). 1970. *Sociological Methods: A Sourcebook*. Butterworths, London

Eggan, Fred. 1954. Social Anthropology and the Method of Controlled Comparison. *American Anthropologist* 56: 743-763

Ellen, R. F. 1984. *Ethnographic Research: A Guide to General Conduct*. Academic Press: London

Fink, Arlene. 1995. *The Survey Hand Book*. Sage Publications: Thousand Oaks

Fisher, R.A. and F. Yates 1982 *Statistical Tables*, Longman: New York

Godde, William J. and Paul K. Hatt 1952. *Methods in Social Research*. McGraw-Hill: New York

Gustavsen, Bjorn 2003. New Forms of Knowledge Production and the Role of Action Research. *Action Research* 1 (October): 153-164

Handel, J.D. 1978, *Statistics for Sociology*, Englewood Cliffs, N.J.

Hughes, John A. and Wes Sharrock 1990. *The Philosophy of Social Research*. Addison Wesley Longman: Essex

- Lazarsfeld, Paul F., Bernard Berelson and Hazel Gaudet 1944. *The People's Choice*. Columbia University Press: Columbia
- Levi-Strauss, C. 1963. *Structural Anthropology*. Penguin
- Lewis, Oscar 1955. Comparisons in Cultural Anthropology. IN W. I. Thomas Jr (ed.) *Current Anthropology*. Chicago University Press: Chicago
- Madan, T. N. 1975. On Living Intimately with Strangers. IN Andre Beteille and T. N. Madan (eds.) *Encounter and Experience: Personal Accounts of Fieldwork*. Vikas: New Delhi
- Madan, T. N. 2004. In Pursuit of Anthropology. IN V. K. Srivastava (ed) *Methodology and Fieldwork*. Oxford University Press: New Delhi
- Mark, Melvin, M. and Gary T Henry 2004. The Mechanisms and Outcomes of Evaluation Influence. *Evaluation* 10:35-37
- Mitchell, J. Clyde 1984. Social Surveys. IN R. F. Ellen (ed) *Ethnographic Research*. Academic press: London. Pp. 271-272
- Moser, C. A. and G. Kalton 1973. *Survey Methods in Social Investigation*. The English Language Book Society: London
- Mukherji, P. N. 2000. *Methodology in Social Research; Dilemmas and Perspectives*. Sage Publications: New Delhi
- Mukherji, P. N. and Chandan Sengupta 2000. Conversations with Ramkrishna Mukherjee. IN P. N. Mukherji (ed.) *Methodology in Social Research; Dilemmas and Perspectives*. Sage Publications: New Delhi
- Murdock, G.P. 1940. The Cross-Cultural Survey. *American Sociological Review* 5: 369-370
- Myrdal, Gunnar 1944. *An American Dilemma: The Negro Problem and Modern Democracy*. Harper: New York
- Nachmias, David and Chava Nachmias 1981. *Research Methods in Social Sciences*. St. Martin Press: New York.
- Neuman, W. Lawrence. 1997. *Social Research Methods*. Allyn and Bacon: Boston
- Reid, William J. 1955. Research Overview, 19th *Encyclopaedia of Social Wor.*, NASW Press: Washington
- Sanders, Donald 1955. *Statistics*. Mc Graw Hill: New York
- Sarantakos, S. 1998 (first published in 1993). *Social Research*. Macmillan: London
- Singleton, Jr Royce A. and Bruce C. Straits 1999. *Approaches to Social Research*. Oxford University Press: New York
- Sjoberg, Gideon and Roger Nett, 1992. *A Methodology for Social Research*. Rawat: Jaipur and New Delhi
- Srivastava, Vinay Kumar (ed.) 2004. *Methodology and Fieldwork*. Oxford University Press: Delhi
- Wallerstein, Immanuel 1997. Social Science and the Quest for Just Society. *American Journal of Sociology* 102 (5): 1241-1257
- Watson, G. and McGraw, D. 1980. *Statistical Inquiry: Elementary Statistics for the Political Science and Policy Sciences*. John Wiley: New York
- Young, P. V. 1988. *Scientific Social Surveys and Research*. Prentice Hall: New Delhi

