

# **Research Methodology**

**(In 2 Volumes)**

**Volume II**

**Suresh C. Sinha**

*Formerly of Central Building Research Institute  
Roorkee - 247 667 (Uttaranchal)*

and

**Anil K. Dhiman**

*Gurukul Kangri University  
Haridwar - 249 404 (Uttaranchal)*

**Ess Ess Publications**

4837/24, Ansari Road, Darya Ganj,  
New Delhi-110002

①: 3260807, 3201533

Fax: 3274173

E-mail: [essess@del3.vsnl.net.in](mailto:essess@del3.vsnl.net.in)

url: <http://www.essessbooks.com>

कृतावाद  
No. \_\_\_\_\_ Date : \_\_\_\_\_  
Name : \_\_\_\_\_ Address : \_\_\_\_\_  
परि. स / लग्न. नं. 6396  
दिनांक / Date : 28-1-05 © Authors

First Published 2002

**ISBN : 81-7000-334-2**

**Price : Rs. 1900/- (Two Vol. Set)**

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of Ess Ess Publications.

*Published by:*

## **Ess Ess Publications**

4837/24, Ansari Road, Darya Ganj,  
New Delhi - 110 002  
Phone : 3260807  
Fax : 3274173  
E-mail : [essess@del3.vsnl.net.in](mailto:essess@del3.vsnl.net.in)  
*url:* <http://www.essessbooks.com>

*Laser Typesetting:*  
**Mukesh Graphics**  
L-2/8, Mangolepuri, New Delhi  
Ph. : 7924405

*Printed at :*  
Print Line

## **Dedication**



**This book is dedicated to  
LORD GANESHA,  
the God of Wisdom and his  
vehicle the Rat, master of Logic**

## PREFACE

The present book on "Research Methodology" has been written keeping in view the requirements of students who are pursuing the Post Graduate studies in the areas of Library and Information Science and social science including sociology, psychology and anthropology of Indian universities leading to the degrees of MLIS and M.A. respectively by various disciplines. It can serve as the main book for them as the topic of the book forms one of the papers in their Masters Degree course. Sinha, the first author had the privilege of offering this paper at M.A. Sociology level in Lucknow University, Lucknow first as PG Student and then as research scholar. In Delhi University, in the Master of Library Science he had another opportunity to study this paper in the year 1970. He wrote 3 dissertations in Sociology, Library Science and at Associateship in Information Science at INSDOC, Delhi during 1977 - 79 under different eminent experts/professors of repute. Sinha's long experience and exposure to work in a Research Laboratory for more than 3 decades had further sharpened his skills, methodology and techniques in the scientific environment.

Dr. Dhiman studied Research Methodology at MLIS and applied the same in writing his thesis which earned him the doctorate in the area of Botany and further sharpened the methodology and this technique got exposure in the library of College of Management Studies at Gurukul Kangri University, Haridwar where he had worked for about four and half years as its In-charge. He has two more dissertations to his credit; One at M.Sc. (Botany) level and another at P.G.D.C.A. (Computers).

Many good books are available in the market written by eminent authors who are teachers of the subject, but most of them are on social research. The present book shall be multifaceted book covering Social Sciences, Sciences and Technology (S&T) and the area of Library & Information Science. Dr. S.R. Ranganathan's approach to library science research and his concept of research on the structure and development of the universe of subjects and his concept of spiral of scientific method is unique as he had combined his mathematical skills in studying the scientific methods. This concept has also been explained.

In compiling this monograph, the authors have banked upon the work of different authors and their ideas have been incorporated. The authors are deeply indebted to them and profusely acknowledge their contributions/works in the course of writing this book. The authors are very much sure that the present book will be useful not only to the students of Library and Information Science leading to their Masters Degree but will be equally useful to all those students who are studying Masters Degree in Social Science viz. Sociology, Anthropology, Psychology as well as the Research Scholars engaged in their Doctoral Degree Programmes

This book is written in two volumes for the convenience of the readers; both of them can be studied independently. First volume of this book covers basic aspects of Research Methodology describing concept of research, research and theory, designing of research, formulation of research and different techniques of research, data analysis, scaling techniques and research report writing etc. Its second volume is devoted to applied aspects of Research Methodology containing chapters on different types of researches such as historical research, descriptive research and experimental methods of research, basic statistical methods and application of computers in researches and application of research techniques in different fields of study.

In the end, the authors can confidently state that this book has a wide scope in the various areas of studies. The authors are grateful to the members of their families, who bore the brunt of the time and their family activities, when the authors were busy in preparing this book. We are indebted and greatful for their moral encouragement and their affection which the authors needed psychologically, and morally. The authors shall feel grateful if a sound criticism is received to further improve the work, by the learned teachers and users of the book. Last but not the least, the authors are also greatful to Shri Sumit Sethi of Ess Ess Publications, who took pains to give this book an attractive form which is in the hands of our readers, the real critics of this book. Wishing our readers a useful reading.

**Suresh C. Sinha**

*Formerly of*  
Central Building Research Institute  
Roorkee - 247 667 (Uttaranchal)

**Dr. Anil K. Dhiman**

*University Library*  
Gurukul Kangri University  
Haridwar - 249 404 (Uttaranchal)

## **ACKNOWLEDGEMENT**

Authors have heavily dwelt upon the the following books, while preparing this book, which is highly acknowledged.

1. Theory and Methods of Social Research, by Galtung, John. London, George Allen and Unwin Ltd., 1967, 954 p.
2. Methods of Research in the Behavioural Sciences by Mc Cormic, Thomas, C. and Francis, Roy. G. New York, Harpar and Brothers, 1958, 244p.
3. Methods in Social Research, by Goode, William J. and Hatt, P.K. New York, McGraw-Hill, 1952. 378 p.
4. Methodology and Techniques of Social Research, by Bhandakar, P.L. and Wilkinson, T. S. Bombay, Himalayan Publishing House, 1996. 458 p.
5. Research Methodology in Social Sciences, by Pandey, G.C. New Delhi, Anmol Publications Pvt. Ltd., 1999. 285 p.
6. Dictionaries and Encyclopedias.

## **CONTENTS**

1	Different Types of Research: Historical Research	... 1
2	Different Types of Research : Descriptive Research	...21
3	Different Types of Research : Experimental Method of Research	...40
4	Different Types of Research : Inter-Disciplinary Research	...57
5	Scientific Methodology	...64
6	Basic Elements of the Scientific and Social Sciences Methods	...79
7	Concepts-Its Role in Scientific Investigation and Development of Science	...121
8	Spiral of Scientific Method	...130
9	Basic Statistical Methods, Concepts and Techniques	...147
10	Computers - Its Application in Research	...250
11	Scientific Social Research	...287
12	Sociometry	...346
13	Research In Library And Information Science	...358
14	Scientometry/Bibliometry	...368

# 1

## **DIFFERENT TYPES OF RESEARCH: HISTORICAL RESEARCH**

It is noted that a scientific inquiry starts with a problem. The problem must be solvable and be started in the form of a question? The next step the formulation of a hypothesis as a possible solution. The hypothesis is verified to determine whether it is true or false. This requires a study in which the empirical results are summarized in the form of an evidence report, of an investigation. Once the evidence report is formed, it is related to the hypothesis. By comparing the hypothesis with the evidence report the hypothesis is tested and data is collected to arrive at an evidence.

### **SCIENTIFIC METHOD IN HISTORICAL RESEARCH**

Man is the only social being who is keenly interested in his past. A.N. White has aptly pointed out; "each emerging is perceived as containing within itself all its past and seeds of its future". The past contains the key to the present. Though today is different from yesterday, it was formed by yesterday. Today and yesterday will probably influence tomorrow. We can hardly conceive of a social situation or a social structure which is not rooted in conditions and forces long in operation. Therefore, Historian Arthur Schlesinger warns; "No individual let alone a social scientist, can wisely ignore the long arm of the past.

## **WHAT IS HISTORY?**

History is an account of the past events. Garraghan writes that the term history stands for three following related but sharply different concepts:

- (a) past human events; past actuality;
- (b) the record of the same; and
- (c) the process or technique or making the record.

This statement implies that the historian deals with the past recorded facts. But the historian is neither the humble slave nor the tyrannical master of facts. The relation between the historian and his facts is one of equality, of give and take. This prompted E.H. Carr to remark that history is a continuous process of interaction between historian and his facts, an unending dialogue between present and the past.

One is further supported by the fact that the Latin word, 'historia', the French 'historie' and the English 'history', originally meant inquiry, investigation, research, and not a record of data accumulated thereby—the usual present-day meaning of the term. It was only at a later period that the Greeks attached to it the meaning of "a record or narration of the results of inquiry." In current usage the term history may accordingly signify or imply any one of three meanings:

(1) Inquiry; (2) the objects of inquiry; and (3) the record of the results of inquiry, corresponding respectively to (c), (a) and (b) of Garraghan's concepts.

History is defined by Garraghan, "as the Science which first investigates and then records, in their causal relations and development, such past human activities as are (i) definite in time and space, (ii) social in nature, and (iii) socially significant."

## **HISTORICAL METHOD**

Every science, when it became an art by being reduced to practice, follows certain rules and directions which insure or help to insure accuracy of result. The complex of these rules and directions

we call as method, or technique. Surgical science has its method of performing a given operation; the musician has a method of handling his instrument; and success in the classroom is very much a matter of effective method. So also with the art of history, the direct aim of which is the attainment of historical truth. This is a complex process involving search for sources of information, critical evaluation of the same, synthesis and exposition of the result of research and criticism. This process must follow certain recognized rules, and regulated by method. Historical method is aptly defined by G.J. Garraghan as a "systematic body of principles and rules designed to aid effectively in gathering the source materials of history, appraising them critically, and presenting a synthesis (generally in written form of the results achieved).

In other words, according to Garraghan, there are three major operations in historical method. First, the search for material on which to work, for sources of information (which he calls heuristic). Second, the appraisement of the material or sources from the viewpoint of evidential value (criticism). Third, formal statement of the findings of heuristic and criticism. This includes the assembling of a body of historical data and their presentation (generally in writing) in terms of objective truth and significance (synthesis and exposition as stated by him). It may, however, be kept in mind that in actual historical work these three operations are not necessarily taken up in secession. They scarcely stand apart (at least the first two); rather they overlap one another, and the historian has synthesis and exposition simultaneously.

## HISTORY AS SCIENCE

Diametrically opposite opinions are held as to whether or not the activities of the historian can be treated scientific or whether there is such thing as historical research. Best and Kahn have offered the following negative and positive opinions:

### NEGATIVE VIEWS

1. Of course, the purpose of science is prediction, the historian cannot usually generalise on the basis of past events. Because

past events were often unplanned or did not develop plans, because there were so many uncontrolled factors, and because the influence of one or a few individuals was so crucial, the same pattern of factors can not be repeated.

2. The historian must depend upon the reported observations of others, often witnesses of doubtful competence, and sometimes of doubtful objectivity.
3. The historian is much like a person trying to complete a complicated jigsaw puzzle with many of the parts missing. On the basis of what is often incomplete evidence, the historian must fill in the gaps by inferring what has happened and why it happened?
4. History does not operate in a closed system such as may be created in the physical science laboratory. The historian cannot control the conditions of observation nor can manipulate the significant variables.

## **POSITIVE VIEWS**

Those who maintain that historical investigation may have the characteristics of scientific research activity present the following arguments:

1. The historian delimits a problem, formulates hypothesis or raises questions to be answered, gathers and analyzes primary data, tests the hypothesis as consistent or inconsistent with the evidence, and formulates generalisations or conclusions.
2. Although the historian may not have witnessed an event or gathered data directly, he or she may have the testimony of a number of witnesses who have observed the event from different points. It is possible that subsequent events have provided additional information not available to contemporary observers. The historian rigorously subjects the evidence to critical analysis in order to establish its authenticity, truthfulness and accuracy.
3. In reaching conclusions, the historian employs principles of probability similar to those used by physical scientists.

4. Although, it is true that the historian cannot control the variables directly, this limitation also characterizes most behavioural research, particularly non-laboratory investigations in sociology, social psychology, and economics.
5. The observations of historians may be described in qualitative or quantitative measurement terms depending upon the subject matter. The absence of quantitative measurement in some historical studies, however, does not prelude the application of scientific methodology, for the application of both quantitative and qualitative descriptions which can be appropriately applied to unscientific investigations.

One can easily see that the precise meaning of the term science is the crux of the whole debate. If the term science necessarily denotes precision or exactness, then history could become a science, only by having its rigorously operation and immutable laws. Plainly speaking with such a conception of science, history is certainly not a science. As is said; "if you can't experiment, if you can't measure, if you can't establish broad unifying hypothesis, if you are not confident in your social engineering, you can't be said to be engaged in scientific study at all." But this is a restricted definition of science.

Speaking broadly, science is a "systematized body of general truths concerning a definite subject matter and established by an efficient (effective) method". This definition given by John F.X. Pyne contains two species; the exact sciences and the non-exact science or social sciences. It recognizes four elements as being essential to the concept of science, whether exact or otherwise:

- (i) **A body of systematized knowledge.** Science is not a heap of isolated, disjointed facts or truths, but a complex of facts or truths closely knit together according to some principle of rational and logical order.
- (ii) **An effective method:** Science is fundamentally a method of dealing with problems. A correct process is one if when applied to, rightly necessarily leads to a correct result. Historian cannot write history without a method. The modern historian

must employ a recognized, correct and consequently effective techniques or method. The use of a recognized method is a prime factor qualifying for writers history to become scientific science.

- (iii) **A definite subject matter:** A scientist works within some more or less sharply defined department or field of human knowledge. The historian, too breaks discipline into many narrow fields which make it possible for the historian to contribute to knowledge on the part by focusing his attention on a manageable subject.
- (iv) **The formulation of general truths.** The general truths or laws established by history are only morally uniform in their application, not rigidly uniform, as in the case of the truths and laws established by the exact sciences.

Obviously, history qualifies as science, if it includes the four elements of *systematized knowledge, effective method, definite subject matter, and general truths*. Its methods of enquiry are critical and objective and the results can be accepted as organized knowledge by a consensus and trained and investigators. The historian thinks of the method of investigation as scientific and the method of presentation of past events as belonging to the realm of history. History brings the natural sciences and artificial sciences into an integrated pattern or process, if we view the gradual advance of man's knowledge as a central theme of historical development. Both historian and scientist examine data, formulate hypothesis and test the hypothesis evidence and reach acceptable conclusions. But compared with natural sciences, history is a science only in broad sense.

## ESSENTIALS OF HISTORICAL RESEARCH

The historical approach is past-oriented and recorded research which appears to illuminate a question of current interest by an intensive study of factual facts that already exists. The better known approach to historical research is that phase of its which seeks to discover material existence of what has not previously known to

recent generation. The good historical researcher, therefore, must be that rare person, expert enough to be able to encompass all of the known information and interrelate it and yet be flexible enough to break the set of pre-existing notions, so that he can see a new relationship explanation, finding, if it exists in data. This makes the historical approach, a particularly difficult one to do well, for in most other fields there is negative correlation between experience and expertness on the one hand and flexibility and set of breaking ability on the other. All this requires a great many things for a researcher if its application is to be useful and correct. Some of the main characteristic on the part of historical researcher may be as under:-

- (1) A great deal of social insight and historical orientations on the part of historical researcher is a must. Events, institutions and opinions are the products of a particular environment at a particular point of time. History is the historian's experience; he who applies the historical method should be able to work out accurately; How? When and Why, the events occurred and evolved the growth of institution. It demands that experiences not only in collecting data but also relating them to the major effective situations and assessing their significance in a particular context. It is this that constitutes social insight which in fact means analysing data in terms of their environment.
- (2) The research in history must take both analytical and critical view of the facts. He must take up an analytical view as chemist analyses a sample in his laboratory. He should also take an over-all-view to assess the true relationships between the events themselves and environments. For example, it is almost impossible to understand the events in the days of the Indian mutiny (1857) or the First War of Independence in Indian history by recording the fights and troops' movements without understanding human emotions which precipitated the crisis.
- (3) The researcher in history should make an objective approach in discovering as well as in interpreting facts. The historian without the facts is worthless and futile; the facts without

their historians are dead and meaningless. History is thus a continuous process of interactions between historian and facts, and unending dialogues between present and the past. Historical method is used in the induction of principles through research to the past social forces which have formed the present. The main objective of this method is to apply mind in the matter of various social problems by discovering the past trends regarding facts, events, attitudes, and by evaluating the ideas of development of thought and action. Under the there is no justification in the criticism of the historian depending upon the reported observations of others, his findings are likely to be of doubtful and sometimes of doubtful objectives. In conclusions, the historian employing the principles of probability similar to those used by the scientists.

- (4) The researcher should be fully conversant with the work already done in his field of research to avoid duplication of efforts. He should also be aware of his own limitations in understanding and interpreting the events as well as the sources he has to collect.
- (5) A great deal of imagination is necessary in using historical method. History cannot be written unless historian can achieve some kind of mental compact of those for whom he is writing.

## STEPS INVOLVED IN A HISTORICAL RESEARCH

### Research Problem

The problem selected for research should have a worthy investigation in itself in addition to its being feasible keeping in view of research ability, time, and cost.

To ensure that their findings are trustworthy historians must rely on primary facts of data. But locating the appropriate sources of data needs imagination, hard-work, patience and resourcefulness. Moreover, the collection of data by historian poses some unique problems. To the data already possibly exist and to generate new data, he can only find them. Similarly, he has no ability to alter the

form in which the data appear but can only accept them in the form in which he finds them.

The historical researcher has no ability to create data, he also has the problem of interpreting these data that is available without asking and clarifying questions. Thus, the historical researcher lacks the basic elements of control necessary the typing data what he wants to produce in the form he desires and accepts it without determining the character of data.

### Sources of Historical Data

David J. Fox classifies the sources of historical data which can be explained as under:

1. Deliberate and Inadvertent Sources.
2. Primary and Secondary Sources.

**Deliberate sources of data** are those in which there is conscious effort to record some event or preserved information. These range from the brief chronological notes on tombstones to paintings or to the extended diaries and autobiographies of historians. While such sources are most often in written form, they also include materials like myths and folk songs which are deliberately passed down from generation to generation by word-of-mouth to keep the material alive.

In contrast are the **inadvertent sources of data** which provide information for the historical researcher even though that was not the original intent of the source. Archaeologists who developed description of cultures based on artifacts they unearth are using inadvertent sources.

Both deliberate and inadvertent sources of information have advantages and disadvantages. The deliberate source of course, undoubtedly presents a subjective view of events. Nevertheless, this source is concerned directly with the event at hand and however subjective, in preserving an individual impression of that event. In contrast, the inadvertent source is being used for something other than it was intended and so the conclusions arrived at must play a

large part in understanding any interpretation of it and the conclusions or inference assured can be right or wrong. But since it is being used for some purpose other than that for which it was created, the inadvertent source must be an objective piece of data. And so, the research must balance virtue against disadvantages.

**Primary and Secondary Sources:** A primary source, according to Fox is firsthand, and always involves the direct reporting or recording of an experience. It can be an *animate primary source*, that is, a person reporting an experience which happened to him or in which he participated. Or it can be an inanimate source, such as in the physical artifacts one finds it in museums; or mechanically reproduced direct records of events or a set of proceedings such films or records; or it can be unrecorded material like the official transcripts or minutes of a meeting or convention. What characterizes both animate and inanimate primary sources is the source of immediacy of experience of the source, and the minimization of intervening persons between the experience and the recording of it. This does not, however, meant that the primary source necessarily is true and accurate in an absolute sense. Anyone who sets down an entry in a diary does not record an impartial, impersonal version of what happened but instead records a personal version with all the possibilities for selective perception, recall consciousness and unconscious distortion. The 'primary data' refers only to the fact that the source is the minimum one step removed from the event itself, it does not refer to the accuracy or truth of the source.

**Secondary sources** are account of an event that were not actually witnessed by the reporter. The reporter may have talked with an actual observer or read an account by an observer, but the testimony is not that of an actual participant or observer. Thus, the secondary sources are not tied directly to the event either as official records or as the memoirs of a participant; these are always at least two steps before the event, occurred.

One can notice two essential differences between primary and secondary sources. First is the difference in immediacy; the primary source is tied directly to the event either as the official

records or as the memories of participant in an event; the secondary sources is not. The second difference is that the secondary source introduce at least one other individual into the production of data, and with this on other individual comes a second a set of possibilities for selective perception, recall, and conscious and unconscious distortion in the furnished data so collected.

**Primary Sources** of data include documents, remains or relics and oral testimony. Best and Kahn have enumerated these as under:

Documents classified as primary sources are constitutions, charters, law court decisions, official minutes and records, autobiographies, letters, diaries, genealogies, census information, contracts, deeds, wills, permits, licences, affidavits, depositions, declarations, proclamations, certificates, bills, receipts, advertisements, maps, diagrams, books, pamphlets, catalogues, films, pictures, paintings, inscriptions, recordings, transcriptions, and research reports.

**Remains or relics** are the objects associated with a dead person, group, or period. Fossils, skeletons, tools, weapons, food, utensils, clothing, buildings, furniture, pictures, paintings, coins, and art objects are examples of those relics and remains that were not deliberately intended for use in transmitting information or as records. However, these sources may provide clear evidence about the past.

**Oral testimony** is the spoken account of a witness of, or participant in an event. This evidence are obtained in a personal interview and may be recorded or transcribed as the witness.

To this list, the more elementary and durable kinds of artistic materials not written in the ordinary sense may be added such as inscription scribed upon clay, chiselled stones, monuments, stamps, coins, woven tapestries, ornamental vessels, scenic or portraits sculptures, historical paintings as in Ajanta and Elora Caves and its portraits. Hand-written materials (*Pandulipis*) including papyrus rolls, bricks bearing cuneiform writings, vellum, parchment, manuscripts, etc. are also primary sources of historical data. So is the case with personal observation.

## EVALUATING HISTORICAL DATA

Evaluating historical data is one of the most complicated problems a historical researcher faces after the collection of India. Evaluation here means analysing and examining the data for its reliability the decision as to which data to believe and which to doubt. It involves the dual process of establishing the authenticity of the source and that of establishing the validity of its content. The process of establishing the authenticity of the data is termed as external criticism and that of establishing the validity of their content is termed an internal criticism.

### EXTERNAL CRITICISM

External criticism checks the authenticity and genuineness of source material. Is the relic or document a true one rather than a forgery, a counterfeit, or a hoax? Best and Kahn suggest various tests for verifying authenticity as:

Establishing the age or authorship of documents signatures may be necessary tests of handwriting, script, type, spelling language usage, documentation, knowledge available at the time and consistency with what is known may involve physical and chemical analysis of ink, paint, paper, parchment, cloth, stone, metals, or wood. Are these elements consistent with known facts about the persons, the knowledge available, and the technology of the period in which the remain or the document originated?

### INTERNAL CRITICISM

After the authenticity of the historical source material has been established, the researcher takes up the internal criticism. It is connected with the validity, creditability, or worth of the thought-content of the document. It also involves such factors as competence, good faith, bias, general reputation, integrity, ability of the author among his contemporaries and the quality of his opinion in the areas under study. Did he actually consider with the facts to reveal a true and clear picture of the project. Did he have any motives for distorting the account? Was he biased under any fear, vanity or pressure? How long after the event tools place he make a record of his testimony,

and was he able to remember accurately what happened? Was he in agreement with other competent witness? These questions are often difficult to answer, but the historian must be sure that the data are authentic and accurate. Only then, the researcher presents them as historical evidence, worthy of serious consideration. But historical evidences are never complete and conclusive and therefore, the conclusion arrived out of it is only a probability.

### INTERPRETING DATA

After the data is collected they are subjected to criticism, then researcher starts the interpretation of the data in relation to the problem. As the unique nature of the historical data, the task of interpretation needs greatest ingenuity and imagination on the part of researcher. He must be very cautious of dealing with 'Cause and effect' relationships. Historical causes are invariably complex and the researcher should know that he is not dealing with clear-cut-cases of causes and effects. George J. Moul aptly comments: "Causation is a troublesome concept in science, it is doubtly so in historical research where 'causes' are in the nature of antecedents and other precipitating factors, rather than 'causes' in restricted scientific sense."

In history, it is difficult to correlate that one event in the chain was caused by the previous event in the chain. Many conditions and circumstances interact and cause a particular event. Therefore, it becomes difficult to assess this accuracy, the influence of a particular event and to identify 'cause and effect' relationships clearly. The researcher should also be cautious in the use of analogy in the interpretation of his data. While making comparisons between one historical event to the other, he must be careful to make use of similarities and differences. Some historical perspective in using evidences and interpretations enable the researcher to evaluate events and personages time and space. His ultimate events and personages distant in terms of time and space. His ultimate determine trends which the data might suggest and draw inferences from the data. His main motive should be to synthesis and interpret rather than addition to the qualities of imagination, resourcefulness and ingenuity.

The researcher must be elegant and objective oriented in style of writing the report. It is extremely difficult task to take often seemingly disparate pieces of information and synthesize them into a meaningful whole. Therefore, he must be permitted a little more freedom on the subjective interpretation of data. Hockett suggests that "the historian is not condemned to a bold, plain, unattractive style" and that "for the sake of relieving the monotony of statement after statement of bare facts, it is permissible, now and then to indulge in a bit of colour". He, however, warns that "above all embellishments must never become a first aim, be allowed to hide or distort the truth".

## **SUMMARY OF THE PROCESS OF HISTORICAL RESEARCH**

The actual process of historical research is summarized by Fox into nine major steps listed below:

1. Determination that the problem selected is appropriate for study through the historical approach.
2. Specification of the population data required.
3. Initial determination that sufficient data are available.
4. Begin data collection through by:
  - (a) Consideration of known data.
  - (b) Seeking latest data from known sources.
    - (i) Primary sources.
    - (ii) Secondary sources
  - (c) Seeking new and previously unknown data.
    - (i) in the form of data.
    - (ii) in the form of sources.
5. Start writing report.
6. Interaction of writing and additional search for examination of data sequence.
7. Completion of descriptive phase of research.
8. Completion of interpretative phase of research.
9. Application of data to present hypothesis for future.

The historical report should be presented in a logical, chronological and topical order-involving consideration of such influences or forces as political institutions, law, economics, geography, social conditions, war, national culture, art, literature, religion, great leaders and natural resources.

Historical research reports should neither be dull and colourless, nor too flowery and ornamental. They must follow precision, continuity of subject, clarity of thought and elegance in their style to give a sense of design and completeness. Obviously, historical research is difficult and demanding.

## LIMITATIONS OF HISTORICAL RESEARCH

### **Dearth of reliable and Adequate Data**

The more older the past, the greater are difficulties of obtaining the relevant facts. Moreover, the researcher does not know whether the data being made available to him are dependable or reliable. He does not know the purpose for which data were collected, by whom it were collected and the extent of subjectivity, if any, in the data.

### **Methods of Keeping Records**

Records are not usually kept in order, so it becomes very difficult for the researcher to trace and locate the requisite records easily and promptly.

### **Dispersal of Documents**

There is hardly any library where the researcher can find all the documents concerning his topic. In the absence of union catalogues, it is very difficult to know the location of desired documents. Historical research requires a careful examination of such documents as court records, records of legislative bodies, letters, diaries, official minutes of organizations, or other primary sources of data preserved by different institutions located far away.

### **Frame of Conditioning Phenomenon**

Data have to be taken into the frame of the conditioning phenomena, otherwise their significance would be lost. Historical

research depends upon the observations made by others. Frame of conditioning phenomena becomes difficult because of the time lapsed between the observation and the time of analysis. The situation under which data were collected cannot be repeated; is an utopian in nature.

### **Verification**

Historical research depends mostly on documents. But neither the data nor the inference drawn can be subjected to verification. It is because such situation to verify the authenticity of data is rather difficult to establish repetition of data correction.

### **Calculation and Measurement of Data: A rare Possibility**

Calculation and measurement of data are not possible in historical research. It is because the researcher cannot be fully aware of the background under which the data were collected. Unless that background is repeated there can be no proper calculations and measurement of data. It is also possible that in measuring the data, some of the important data might be missed or underestimated and less important data might be overestimated.

According to Paul V. Young, limitations also emerge in the writing of history because (a) historians cannot write life-size history; (b) not all happenings in time and space can be known at time of writing; (c) personal biases and subjective interpretations often enter unconsciously, even when honest, attempts are made to select pertinent facts, to arrange them consistently, and to place them in a coherent and true frame of reference.

It is not possible to estimate the period as it would require long time to complete any specific project. This is particularly true if the research problem demands a search for new data. But even in simpler historical problems based on reworking or reintegration of existing data, it is difficult to predict a completion date. In fact, any significant, historical study would make demands that few students have the time, financial resources, patience or expertise to meet. That is why, good historical studies are not often attempted.

## Sampling

Then there is a problem of selecting sample. History belonging to a particular period, institution, organisation, society, dynasty, empire or nation is very vast than *prima-facie* it appears. It becomes a life-long work to examine beliefs, behaviours, patterns, situations and institutions. For these reasons, no historian has either patience, time or resources to study all or some of many narrow fields which make possible for the historian to contribute to knowledge of the past by focusing his attention on a manageable subject. That is a way of sampling. By breaking man's history into small, periods historian can obtain the general knowledge about a period that is the foundation of original research. The method of sampling would enable the historian to gain considerable excess to the knowledge of socio-economic, religious, political and intellectual conditions of a particular period. In the words of Paul V. Young historical data may be regarded as adequate and reliable for social research (i) when they are presented as complexes of social forces: (ii) when social phenomena meaningfully depict intricate social processes; and (iii) when sets of inter-relationships, psychological, economic, educational, political and religious-contribute to a unified whole, a configuration or complex pattern.

## Merits and Demerits of Historical Research

Notwithstanding its limitations, the historical method is recognised as one of the important methods for the study of social sciences. Rather there are some social problem, which can be investigated only by this approach. Many a times series data are used for assessing the progress of the impact of several policies framed for social welfare and this is possible by pursuing the historical records only. Historical data thus provide diversified information for evaluating the strength and weakness of a theory as well as many fruitful clues for understanding many unknown social problems. It provides a basis for measuring factors, in the formation of social theory. This method helps us to know how a particular idea was conceived and

if the experiment failed, what were its causes. For such problems only historical approach would suit better.

### **Summary and Conclusions**

Historical research is the application of scientific method to the description and analysis of past events. The historical methods involve a procedure supplementary to observation, a process by which the historian seeks to test the truthfulness of the reports of observations made by others. It implies the use of records of precious past events with the aim of arriving at generalisation that may be used for the solution of current problems by means of discovering past trends of events, facts and attitudes. It traces lines of development in human thought and action in order to reach some basis for social activities. All history according to Collingwood, is the history of thought. Historical method, therefore, is the systematic evaluation and synthesis of evidence in order to establish facts and draw conclusion concerning the past. As Paulin V. Young has put it; "historical data drawn from a variety of sources, judiciously chosen, critically examined and discriminately used, may constitute a fund of knowledge indispensable in understanding and generalising on community phenomena, the social milieu and social institutions." A successful application of the historical method requires careful tolerant, sympathetic but analytical and critical attitude on part of the researcher.

### **HISTORICAL RESEARCH AT A GLANCE**

Historical research is the critical investigation of events, developments, and experiences of the past, the careful weighting of evidence of the validity of sources of information on the past, and the interpretation of the weighted evidence. The historical investigator, like other investigators, then, collects data, evaluates the data for validity, and interprets the data. Actually the historical method, differs from other scholarly activity only in its rather elusive subject matter, the past, and the peculiarly difficult interpretative task imposed by the elusive nature of its subject matter.

One of the basic rules of research in history is; always use primary sources. A primary sources is the original repository of an

historical datum, like an original record kept of an important occasion, an eyewitness description of an event, a photography, minutes of organization meetings, and so on. A secondary source is an account or record of an historical event or circumstance one or more steps removed from an original repository. Instead of the minutes of an organization meeting, for example, one uses a newspaper account of the meeting. Instead of studying and citing the original report of a research, one studies and cites someone else's account and digest of it.

To use secondary sources of data when primary sources are available is a major historiographical error. Materials and data, especially those about human beings and their activities, are changed and often distorted in transmission. The reputable historian never completely trusts over secondary sources, though, he, of course, studies them and weighs them for their validity. The dangers of distortion and consequent erroneous interpretation are too great.

Two other canons of historical method are expressed by the terms external criticism and internal criticism. The historian critically examines the sources of data for their genuineness, or more accurately, for their validity. Is the document or source genuine? If X wrote the paper, was he competent and truthful witness to write such a paper. This is external criticism, internal criticism is preoccupied with the content of the source or document and its meaning. Do the statements make accurate representations of the historical facts? A document may survive external criticism and still be suspected as evidence. There may be no doubt of the "true" author or recorder of events and he may be competent. Willingly or unwillingly, however, he might have distorted the truth. Internal criticism, in brief, seek the "true" meaning and value of the content of sources of data.

The following except from a report of a Committee of Historians on Historiography summarizes the importance of Historiography to the social sciences.

Historiography has a necessary relevance to all the social sciences, to the humanities and to the formulation of public and

private policies, because (1) all the data utilized in the social sciences, in the humanities, and in the formulation of public and private policies are drawn from records of, experience in or writing about the past and because (2) All policies respecting human affairs, public or private, and all generalizations of a nonstatistical character in the social sciences and in the humanities involve interpretations of or assumptions about the past and because workers in the social sciences and in the humanities are personalities of given times, places, and some measure conditioned and determined by the historical circumstances of their lives and experiences.

માત્ર માનુષ કાળજી મંદિર, સિક્કારાવાડી  
National Institute for the Mentally  
Handicapped, Sikka Ravada  
પાઠ. નં | Acc. No. 6396  
દિનાં | Date : 28-1-2005

## 2

# DIFFERENT TYPES OF RESEARCH : DESCRIPTIVE RESEARCH

A descriptive study is present-oriented research. It describes and interprets what it is about? It is primarily concerned with conditions or relationships that exist, opinions that are held, processes that are going on, effects that are evident, or trends that are developing.

Descriptive research deals with the relationships between variables, the testing of hypothesis, and the development of generalisations, principles, or theories that have universal validity. In other words, the researchers engaged in descriptive studies carry on the same activities as researchers of other fields do: They (i) Identify and define their problem (ii) List the assumptions upon which their hypothesis, and procedures are based (iii) state their objectives' hypothesis (iv) choose appropriate subjects and source materials (v) select or construct tools for data collection (vi) specify categories of data that are relevant for the purpose, and capable of bringing out significant similarities, differences, or relationships (vii) describe, analyse and interpret their data in clear and precise terms and (viii) draw significant and meaningful conclusions.

Descriptive research differs from other types of research in purpose and scope. A clear-cut distinction is visible between descriptive

studies and historical studies on the basis of time. The historical studies discover, describe and interpret what existed in the past. Descriptive research, in contrast, is present-oriented. The limitations of both these types of research are similar with regard to the 'cause and effect' relationships that are difficult to establish, and the time at which study is conducted is a critical factor in the interpretation of the data. But in contrast to an experimental research, the descriptive research, is relatively less scientifically sophisticated. In it, the researcher does not manipulate the variables or arrange the events to happen. Descriptive studies involve events that have already taken place. In fact, the events that are observed and described by him would have happened even though there had been no observation.

Descriptive research differs greatly in complexity, at one extreme, it constitutes nothing more than frequency count of events to the study of local problems without any significant research purpose. On the other extreme, it attempts to ascertain significant inter-relationships among phenomena.

The descriptive method is extensively used in social sciences—socio-economic surveys, job and activity analysis, motivation, and expectations studies. The objective of socio-economic institutions, area or group with a view to infer generalisations helpful to solve the problem directly or to guide other investigations. Job analysis takes note of the man-power available and the man-power required for different jobs in various professions. It is to be noted that the complex theoretical and philosophical problems in social sciences cannot be subjected to descriptive method. The problem must be capable of being described and not merely argued about.

In addition to social sciences, the descriptive method is also used in physical and natural sciences, for example, when physics measures, biology classifies, zoology dissects and geology studies the rocks.

The descriptive research is useful in the development of data collection instruments and tools like checklists, schedules, score questionnaires, opinion surveys and rating scale. It also provides

the background ideas and data from which many more refined or controlled studies of causal relations are made.

## CHARACTERISTICS

The descriptive research studies have certain characteristics:

1. They involve hypothesis formulations and testing.
2. They use a logical method of inductive-deductive reasoning to arrive at generalizations .
3. They often employ methods of randomisation so that error may be estimated when inferring population characteristics from observations are sampled.
4. The variables and procedures are described as accurate and complete as possible so that the study can be replicated by other researchers.
5. They are non-experimental, for they deal with the relationships between non-manipulated variables in a natural rather than artificial setting. Since the events or conditions already exist, the researcher select the relevant variables for an analysis of their relationships.

## TYPES

The descriptive studies may be classified in the following four categories:

1. Survey studies
2. Case study
3. Developmental studies and
4. Content analysis

### Survey Studies

As the survey is present-oriented research, it is suitable for problems in which the researcher believes that although the data needed to resolve his research question do not exist, the settings in which those data could be generated do exist. The researcher's approach then is to go to those settings, administer appropriate data collection devices and analyse the data. In the sense that the settings

are known and the data required reasonably well-defined by the statement of the research problem, the survey approach is more structured for the researcher than the historical approach. However, surveys have their own problems; the difficult problems of instrument development-data collecting devices.

### **Surveys Research**

Many research problems require systematic collection of data from population through the use of personal interviews or other data gathering devices. These studies are usually called surveys, especially when they are concerned with large and widely diverse groups of people.

- \* A social survey is a process by which quantitative facts are collected about the social aspects of a community's composition and activities.
- \* A social survey is a collection of data concerning the living and working conditions of the people in a given society.

### **Purpose of Surveys**

The purpose of most of the surveys is simply to provide someone information. That someone may be a government wanting to know how much people spend on food; business concern interested to find out what detergents people are using; a research institution studying the housing of old age pensioners and so on. The surveys are conducted to collect detailed descriptions of existing phenomena with the intent of employing data to justify current conditions and practices or to make more intelligent plans for improving them. Their objective is not only to analyse, interpret, and report the status of an institution, group, or an area in order to guide practice in the immediate future, but also to determine the adequacy of status by comparing it with established standard. Some surveys are confined to gather all three types of information viz; (i) data concerning existing status, (ii) comparison of existing status with the established status and standards, and (iii) means of improving the existing status; while others are limited to one or two of these types.

A full inventory of all types of surveys research is out of place here but we may take note of three types of survey given by Fox stemming from the researcher's intent to describe, compare or evaluate. These types are the descriptive survey, concerned with describing a specific set of phenomena at one point in time; the comparative survey, intended to compare two or more research situations in terms of some preselected criteria; and the evaluative survey, planned to evaluate some aspect of a research situation, again in terms of preselected criteria.

### **Characteristics**

The survey has the advantage of being an extremely effective way of gathering information from a large number of sources relatively cheap and in relatively short time. Facts once gathered through the use of questionnaires, interviews, standardized tests and other data gathering techniques and the analysis of such information has enabled decisions to be made which have transformed many administrative, financial and other practices.

- \* Survey study is essentially cross-sectional, mostly of what exists type, that is to say it is designed to determine the nature of the existing state of affairs.
- \* It is an important type of research involving clearly defined problems and definite objectives. It requires an imaginative planning, a careful analysis and interpretation of data and a logical and skilful reporting of the findings.
- \* It does not aspire to develop an organised body of scientific laws but provide information useful to the solution of local problems. It may, however, provide data to form the basis of research of a more fundamental nature.
- \* Survey vary greatly in complexity, some concerning themselves with the frequency count of events, while others to establish relationship among events.
- \* Surveys may be qualitative or quantitative. At one level survey or status studies may consist of naming and defining constituents,

elements of various phenomena. At another level, they may involve the amount of constituents or characteristics.

Descriptions may be either verbal or expressed in mathematical symbols.

It fits appropriately to the total research scheme or the expressed stages in exploring a field of investigation. It may serve as a reconnaissance or getting acquainted with the stage of research in entering a new area. Or it may represent a specific interest in current conditions within a field that has long since been explored and developed by research.

### **THE PROCESS OF SURVEY RESEARCH**

1. The research problem must be clearly identified and stated precisely. This must be based on the research previously carried out to investigate the problem.
2. The nature of the sample to be studied must be determined.
3. Selection of the appropriate type of survey. The research instruments to be used must be specified.
4. Design of data, Collection of data and Analysis of data.
5. Preparation of the report, Descriptive phase, Comparative phase and conclusions. The result must be disseminated in terms which are precise and unequivocal.

The survey is an important type of study. It must not be confused with the mere clerical routine of gathering and tabulating data. It involves a clearly defined problem and definite objectives. It requires expert and imaginative planning, careful analysis, and interpretation of the data gathered, and logical and skilful reporting of the findings. In Library's Glossary it is defined as "a scientifically conducted study through which data are gathered according to a definite schedule and are presented in statistical, tabulated or summarized form".

### **MERITS OF SURVEY RESEARCH**

Survey research has come to stay in modern society. Although

this type of research is time-consuming and also costly, yet if conducted consciously with care it has its own merits.

Surveys help us to scrutinise a problem in all its aspects thoroughly and deeply. It is very useful for administrators and policy-makers. Social Surveys help us in finding the trends of changes as well as disorganisation in society. Since an investigator comes vis-a-vis in contact with respondents, about whom he wants to know, the data collected are based on the realities of life, therefore, these are dependable and authentic. Moreover, as a social survey is conducted by a large number of people, the chances of subjectivity are also lessened. Surveys highlight the hidden problems. The comparative survey helps us to study the socio-economic or political conditions of top societies.

### **DEMERITS OF SURVEY RESEARCH**

The demerits of survey research are not less potent. First, survey requires heavy funds and trained investigators, field workers, and researchers. It is difficult to marshal the requisite economic and human resources. Similarly, at the administrative level, people are requested for coordinating resources that will enable accomplishment of specified ends or goals. Secondly, when a large number of people are sent to the field, there can be no uniformity in the collection of data. Everyone engaged in the collection of data may give his own interpretation to the terms used in the data collecting device and where there is no uniformly, the results are bound to be variant. Moreover, everyone engaged in data collection may laad his own bias in the study. It becomes very difficult to find out the extent of bias and much efforts are needed to minimise the bias so that its effect on net result becomes negligible. Thirdly, to have dependable and reliable data as far as possible, two conditions have to be met: (i) trained investigators with a sense of responsibility, integrity and dedication to duty must be engaged; and (ii) the respondents must be cooperative. But both these pre-requisites are difficult to attain. Fourthly, survey method of research is time-consuming, and people do not have time to wait for long time in modern life. If a survey is conducted in a hurry, we know the proverb that hurry

spoils curvy; much will be left out and whole survey will become undependable and unreliable. During the course of survey, there is temptation to collect maximum information. After putting up much efforts, it is not found useful to analyse and tabulate all the information. But one also does not want to destroy the unused information. This leads to the problem of information storage and retrieval. Survey is either to generalise or to localise. If survey is conducted at a local area, then, the results would be applicable to that area only and cannot be generalised. If survey is spread to a vast area, then the results are too general to be of much use. Thus, in both the cases, the survey has its own limitations. Finally, survey of the entire population cannot be undertaken. The way out is to pick up representative sample of the population. But when we use sampling method, then sampling errors are bound to creep in.

It will be profitable for the students of library and information science to have a bird's eye-view of library surveys.

## **LIBRARY SURVEY**

The Librarian's Glossary defines the survey as "an account of some research, examination , or enquiry which has been done by a scientific or organized method". The goal of library survey is the improvement of library services. It is a specialised type of investigation. The Surveyor evaluates library programme and operations and makes suggestions for improvements.

Library survey is a systematic collection of data, their programmes and activities, operations, staff use and user at a given time or period. A study of a whole library and its programmes and operations is called a survey. Cataloguing costs and services within several libraries is a survey. Similarly, the comparison of reference services in large number of libraries is a survey. It is clear that survey is not possible within few figures, but it requires heavy statistics for analysis, so that the surveyor may be able to find out the result. A survey gives an oversight of a field, and is thus distinguished from the sort of study which consists of a microscope examination of the turf; it is a map rather than a detailed plan.

A common but complex element of survey is measurement. Some valuable services provided by the library cannot truly be measured. A surveyor cannot measure the value of a library to its community. Even a surveyor cannot measure exactly the use made of library by the community for its elevation. Behavior of the community and economic contribution by the community is also unmeasurable. These types of values are not exactly measured but there is only judgement.

Surveys are often classified into descriptive and analytical. The descriptive survey is content to enumerate and describe. It is a compendium of concise and comprehensive account of information, but does attempt to do anything very clear with the data collected. The analytical survey does rest content with collecting and arranging data. It attempts to relate one piece of data to another, to probe beneath the figures, the underlying factors and patterns. But the division of surveys into descriptive and analytical is rather unreal. If there is a valid distinction, it is in the aims rather than in methods. The descriptive surveys are library oriented, whereas 'analytical' surveys are subject-oriented.

## OBJECTIVES OF LIBRARY SURVEY

The objectives of survey of library collection are:

- \* to examine and evaluate the nature of collection for an implied or specified purpose;
- \* to determine the condition of collection regarding its internal organisation and maintenance and its relevancy to the aims and objectives of the parent organisation to which the library is attached;
- \* to make known, the strength of general libraries in special subjects;
- \* to discover unsuspected or little known collections of real importance;
- \* to acquaint scholars and other research workers with the collections likely to be most useful to them;

- \* to assist in spreading library use, with possible relief to some large libraries;
- \* to assist in spreading inter-library lending;
- \* to supplement permanently the use of a Union Catalogue, both current and retrospective types;
- \* to estimate the value of a collection in comparison with the collection of some other library of similar nature;
- \* to make a comprehensive or selective study of a specified subject field or group of several fields with an intention to draw some helpful conclusions, e.g. what impact on library use does publicity have; and
- \* to present the results of evaluation or study in the form of special document or report.

## **USEFULNESS OF LIBRARY SURVEY**

Library surveys are very useful for library staff, faculty, research scholars, administration and other users. Some of the uses may be stated as under:

1. Library surveys are used for references in book-selection;
2. these are used in connection with the study of a particular library and other sources of the library;
3. these are also used as a reference manual for a research scholar or reference librarian;
4. the teaching staff uses library surveys to keep up with current trends in administration and technical processes;
5. they are used to illustrate particular aspects of library problems;
6. these are used as a useful source to determine the relative strength and weaknesses of collections.
7. they provide that may lead to generalization of certain problems in respect to solutions; and
8. these can also be used to demonstrate the condition of the library to higher authorities. Its collection and how well the

library is serving its clientele. Surveys are also used to measure user's demands and requirements.

The value and effectiveness of library survey depends on the competence of the surveyor and the willingness of the librarian, library staff, faculty and administration to follow the advice of the expert. Wherever conducted properly, the library surveys have contributed strongly to the improvement of library service and status of the profession. It has provided authors with some of the essential source material on major library problems, i.e. finance, personnel administration and organisation. It exerts a broadening influence on library services by sharing the impact of libraries on society. Probably, it has made more contributions to individual library progress and improvement.

The survey research has contributed to the movement towards a scientific evaluation of libraries and their services. It starts with questionnaire and statistical tabulation. It uses and tests most of the surgical tools of the research process. It shows clearly the value of standards, test them and points the way towards their revision. It helps to teach pragmatic librarians about theoretical concepts of library government, book selection policies, and administrative principles. Usually library survey deals with many problems in combination or with any single isolated item.

Survey techniques of today are as much pretty as the same were in use a decade ago. The under developed techniques of survey has been a problem in applying scientific methodology.

## II. CASE STUDIES

A case study is not different from a survey. But instead of collecting information about few factors from a large number of units, the researcher makes a depth and intensive study of a limited number of representative cases. It is narrower in scope but more exhaustive and more informative in nature than a survey. To provide more qualitative data, the case study is often used to supplement the survey method.

Some definitions of case study method of research are:

- (i) The case study is a way of organizing social data so as to preserve the unitary character of social object being studied. Expressed somewhat differently, it is an approach which views any social unit as a whole.
- (ii) Case study is a method of exploring and analysing the life of a social unit, be that a person, or family, or institution, cultural group or even entire community.
- (iii) The case study is a form of qualitative analysis involving very careful and complete observation of a person, or situation, or of an institution.

Case study aims at an inclusive and intensive study of the unit. Unit of study may be an individual, a family, an institution, a cultural group or a community, the whole nation, empire or historical age, or set of relationships or processes (such as family crises, adjustment to disease, friendship formation, ethic invasion of a neighborhood, etc.), or even an entire culture. Case studies have been made of all types of communities, and all types of individuals.

The basic rationale for the case study is that there are processes and interactions, such as aspects of personality and social functioning which cannot be studied except as they interact and operate within an individual. Moreover, the probability is that if we learn how these processes interact in some individuals we shall also learn much about the processes in the abstract, and ultimately learn all there is to know about them.

In a case study, data may be collected by a wide variety of methods, including:-

1. Observation by the researcher or his informants of physical characteristics, social qualities, or behaviour.
2. Interviews with the subject (s), relatives, friends, teachers and counsellors etc.
3. Questionnaires, opinionnaires, psychological tests and inventories.
4. Recorded data from newspapers, schools, courts, clinics, government agencies, or other sources.

As already stated, in a case study the researcher attempts to examine an individual or unit in depth. He gathers pertinent data about the present status, past experiences, and environmental forces that contribute to the behaviour of the individual or social unit, and how these factors relate to one another. The analysis of the factors and their interrelationships, helps the researcher to construct a comprehensive and integrated picture of the unit.

Case studies of individuals are carried out in the following two ways:-

### **Clinical Case Study**

It is usually carried out by Psychiatrist, artists, and social workers, in order to diagnose a particular condition with a view to recommending therapeutic measures. The individual is studied as a unique personality.

### **Biographical Case Study**

An account of the individual is provided by means of prolonged study and numerous observations which may spread over years. Whole life span is rarely studied, the study is usually confined to some part of life.

### **Merits of Case Study**

Case study method of research is used by trained and experienced persons and it reveals data in a way provided by no other form of research. It is an extremely useful method to study rare or unusual cases of human behaviour, particularly in the study of the handicapped subjects and those suffering from rare physical conditions.

Moreover, the case study attempts to understand an individual (or a single incident of his life) in the totality of his environment and thus to have deep probing-the sheet-anchor of research. The researcher develops insights into basic aspects of human behaviour; that may help him to formulate fruitful hypothesis of a set of hypothesis.

Case study is helpful in framing questionnaire and in stratification

of the sample. It is also possible to locate deviant cases-those units that behave against the proposed hypothesis. The analysis of such cases may lead to a lot of clarification of the theory itself.

### **Demerits of Case Study**

There is a danger of subjectivity because there is too much association of the researcher with the subject of study. The elements of subjectivity enter into report, particularly when judgements are made about the subject's characters and motives. The researcher's personal biases and standards, may influence his interpretation.

Of course, case study method attempts to examine an individual in depth, it inevitably lacks breath. Most case studies generally arise out of counselling or remedial efforts and provide information of exceptional rather than representative individuals. Moreover, it is impossible to either confirm or refute through empirical study the findings and results of a particular case study. Then, the researches based on the study of a single case or a few cases are extremely expensive in terms of time and financial cost and it may only be justifiable in relation to cases which are unusual or exceptional. Also, a worthwhile case study can rarely be completed by a single individual. It usually requires a team of experts. Biographical documents do not provide impersonal, universal, non-ethical, non-practical, repetitive aspect of phenomena. Case situations are seldom comparable.

The demerits of the case study have not discouraged the social scientists to drop the method as unscientific, unsystematic and unfit for valid generalisations. Rather they have made concerned efforts to put the method on scientific lines. Attempts have been made to quantify the case data and thus render it capable of statistical analysis; various kinds of socio-metric scale need a special mention in this respect. Despite the criticism, the case study remains one of the most fundamental methods of study of the social phenomena.

## **III. DEVELOPMENTAL STUDIES**

Developmental studies are used for investigating the characteristics of children and the ways in which these characteristics change with growth and development. It is essential for the

educationists to have reliable information about physiological, intellectual and emotional growth of children at various ages, how they differ from one another within certain age levels, and how they change as a result of certain treatments. Developmental studies are concerned not only with the present status and interrelationships of phenomena but also with the changes that take place as a function of time. Such studies are also called as genetic studies. The genetic or developmental studies may be of following types:

- (i) growth studies; (ii) follow-up studies; and (iii) trend studies.

### **Growth studies**

The growth studies are used to seek knowledge of the nature and rate of changes that take place in human organism. These are helpful to watch when various aspects of growth among children are first observe, accelerate further, remain rather stationary, attain optimal development; and decline.

Growth studies can be either longitudinal or cross-sectional.

Latitudinal studies follow the same subject, a group of subjects, or a situation, or any institution over a relatively long period of time. It is a kind of extended case-study as it involves a long-life span of the same subject, group or institution. A researcher, for example may test and measure the same student on a trait of personality when he is eleven, twelve, thirteen, fourteen and fifteen years of age and plot his individual growth patterns for this trait for personality during these years. This type of research is mostly used in clinics and laboratories.

An entirely different approach to evaluating change over time is to use the Cross-sectional study, which evaluates change over time by comparing at the same point in time different people representing different stages of development instead of following the same group of individuals and taking their repeated measurements over a relatively long period of time as in longitudinal studies. In cross-sectional studies, the random samples of subjects of successive ages are selected and one set of measurements of different individuals from each age level are taken as the basis for developing growth

norms. The statistics derived from the samples are compared and the conclusions are drawn about the growth of individual with respect to the variables studied.

Comparing longitudinal and cross-sectional studies, one can finds that each has its own strengths and weaknesses. The major strengths of the cross-sectional design are that it provides the researcher with his data quickly, eliminates the problem of meeting the same respondents at different points in time, and by using different people to represent the different time intervals avoids any possibility that the data, in part, reflect the test-witnesses of the respondents. Its great weakness is that it rests on the assumption that each element at an advanced point in time provides us with the same data that earlier elements would, if we waited and did a longitudinal study.

The great advantage of the longitudinal design is that it avoids the assumption of comparability of different groups by using the same respondents at every data-collection interval. Its weaknesses are the length of time required to obtain the complete set of data and the difficulty in maintaining contact with, and assembling, the respondents throughout the length of the study. Moreover, continued exposure to the data-collection instruments and the research produces a degree of sophistication or test-wiseness in the respondents which is rejected in the data.

The latitudinal studies are generally more acceptable because they have the merit of continuity.

**Follow-up-studies** aim at investigating the subsequent development of individual (s) after a specified treatment or condition. In case of study approach, a researcher may by interested to follow-up a case in order to determine whether the treatment is successful. It is found unsatisfactory, a new diagnosis of the problem(s) is made. The techniques of experimentation are employed to evaluate the success or condition.

**Trend Studies** are used to obtain and analyse social, economic, or political data to identify trends and to predict what is likely to take place in the future. Trends studies are conducted through

documentary analysis or surveys at repeated intervals. Sometimes trend study may combined with the historical, documentary, and survey techniques.

#### **IV. CONTENT ANALYSIS**

It is an important type of descriptive research which is also referred as documentary analysis. The analysis is concerned with the explanation of the status of same phenomenon at a particular time or its development over a period of time. The activity may be classified as descriptive research, for problem identification, hypothesis formulation, sampling, and systematic observation of variable relationship may lead to generalisations. It serves a useful purpose in adding knowledge to the field of inquiry and in explaining certain social events. The methods of analysis are similar to those used by historians.

In a documentary analysis, the followings are usually used as sources of data: records, reports, printed forms, letters, autobiographies, diaries, compositions, books, periodicals, bulletins, catalogues, syllabi, court decisions, pictures, films and cartoons. Documents used in descriptive research must be subjected to the same careful types of criticism employed by the historians. Not only is the authenticity of the document important, but also the validity of its contents is crucial. It is the researcher's obligation to establish the trust worthiness of all data he draws from documentary sources.

#### **Merits of Descriptive Research**

By and large, the scientists do not treat the descriptive research as research of high order. But the type of information procured by this method is in wide demand and is capable of rendering important service due to the following reasons:

The existing theories and laboratory findings can easily be put to test in real situations. Many times, revealing facts come to light which may substantiate the existing doctrines, lead to their improvements of modifications or may necessitate the changes in their assumptions and help in building a new theory altogether.

Sometimes, laboratories conduct experiments and draw certain conclusion. Descriptive method seeks to verify these conclusions in field investigations. The laboratory findings may hold good or may have to be modified in the light of field investigation. It provides the background ideas and data from which many more refined laboratory or controlled studies of causal relations are made.

It determines the present trends and solves the current practical problems. It also secures historical perspective through a series of cross-sectional pictures of similar conditions at different times. It also suggests the course of future developments.

This method brings the researcher and the respondent face to face and their cooperative efforts help to build up a better research study. It also helps in engineering various data collection devices.

### **Demerits of Descriptive Research**

Descriptive research is a fact-finding approach but quite often it becomes a data gathering process. Merely fact-finding is not a research. The research involves discovery of facts and it is essentially creative. Fact collecting is no substitute for thought and content analysis. Good judgement sometimes requires deliberate sacrifice of quantitative precision for the greater depth attainable by more intensive inquiry and by the application of sophisticated research methodology. Another limitation flows from the desire to over emphasise statistics-central tendencies, and to present the facts in terms of average, co-efficient of correlation, mean, median, mode and standard deviations. Statistics is a descriptive tool; it is not always helpful to find out causal relations in an accurate manner.

Descriptive research involves considerable time and efforts to complete the investigations. Sometimes data collection process takes too long time to complete investigations and the data become out-dated by the time. In such cases many human-hours's labour is wasted without contributing to the advancement of knowledge. Moreover, much depends on the cooperation of the respondents over whom the researcher has no control. The problem becomes more acute when dealing with illiterate and ignorant respondents

because of their indifference, inhibitions and lack of awareness of civic and academic investigation.

## CONCLUSION

Limitations, however, do not deter the social scientists particularly in developing countries to apply this method to a number of social problems. It is essentially a fact-finding approach related largely to the present, and drawing generalisations through the cross-sectional study of the present situation. This research method is useful in identifying objectives and to highlight the ways for attaining them. This mainly consists of data- collecting approach. But since mere collection of data does not constitute research, the data have to be properly interpreted to find causal relationships. Finally, to apply the descriptive method, the data collected should be accurate, objective and, if possible, quantifiable. This will make the data more satisfactory and reliable. One should also ensure that the descriptions are objective, adequate and accurate.

# **3**

## **DIFFERENT TYPES OF RESEARCH : EXPERIMENTAL METHOD OF RESEARCH**

Experimental research is a most sophisticated, exacting, and powerful method for discovering and developing the organized body of knowledge. Experimentation is the classic method of science laboratory, where elements manipulated certain stimuli, treatments or environmental conditions and observe how the condition or behaviour of the subject is affected or changed. Their manipulation is deliberate and systematic.

Experimentation provides a method of hypothesis testing. After experimenters define the problem, they propose a tentative answer, or hypothesis. They test the hypothesis and confirm or deconfirm it in the light of the controlled variable-relationship that they observed. It is important to note that the confirmation or rejection of the hypothesis is stated in terms of probability rather than certainty. From this description follow the definitions of experimental research.

1. The essence of an experiment may be described as observing the effect on a dependent variable of the manipulation of an independent variable (Festinger).
2. In experimentation the investigator controls (manipulates or

changes) certain independent variables and observes the changes which take place in the form of dependent variables. (C.V. Good).

3. An experiment is the proof of a hypothesis which seeks to look up two factors into a causal relationship through the study of contrasting situations which have been controlled on all factors except the one of the interest, the latter being either the hypothetical cause or the hypothetical effect (E. Greenwood).
4. The term experiment refers to that part of research in which some variables are controlled while others are manipulated and their effects on controlled variables are observed (Vimal Shah).
5. Experiment is simply observation under controlled conditions. When observation alone fails to disclose the factors that operate in a given problem, it is necessary for the scientist to resort to experiment (F.S. Chaplin).
6. In its broadest meaning, an experiment may be considered as a way of organising the collection of evidence so as to permit one to make inferences about tenability of a hypothesis (Jahoda and Cook).

The essence of all these definitions is that the experimentation is an observation and a way of organisation of the data. Experimental studies are concerned with testing the causal hypothesis. This method helps the experimenter to test the reality of the problem under controlled conditions or variables.

## VARIABLES DEFINED

Variables are the conditions or characteristics that the experimenter manipulates, controls, or observes. The independent variables are the conditions or characteristic that the experimenter manipulates or controls in his or her attempt to ascertain their relationship to observed phenomena. The dependent variables are the conditions or characteristics that appear, disappear, or change as the experimenter introduces, removes and changes independent

variables. For example, in educational research, an independent variable may be a particular teaching method, a type of teaching material, a reward, or a period of exposure to a particular condition, or an attribute such as sex or level of intelligence. The dependent variable may be a test score, the number of errors, or measured speed in performing a task. Thus, the dependent variables are the measured changes in pupil performance attributable to the influence of the independent variables.

Independent variables are of two types; **treatment variables** and **organismic or attribute variables**. The former are those factors that the researcher manipulates and to which he assigns subjects. The latter are those characteristics that cannot be altered by the researcher, e.g. age, sex, race and intelligence level etc.

Then we come across confounding variables. These variables are those aspects of a study or sample that might influence the dependent variable (outcome measure) and whose effect may be confused with the effects of the independent variable. Confounding variables are of two types; intervening and extraneous variables.

**Intervening variables.** In many types of behavioural research the relationship between independent and dependent variables is not one of stimulus to response. Certain variables (e.g. anxiety, fatigue, and motivation) which cannot be controlled or measured directly may have an important effect upon the outcome. These modifying variables intervene between the cause and effect.

**Extraneous Variables** are those uncontrolled variable (i.e. variables not manipulated by the experimenter) that may have a significant influence upon the results of a study. Many research conclusions are questionable because of the influence of these extraneous variables. These can, however, be controlled by randomization, matching case and balancing case or group matching, etc. as we shall study.

The basic factor in case of experimental research is the control over the subject of study and manipulation of the independent variable to study its effect upon the dependent variables. In the

narrowest sense, only pure laboratory experiments are to be included in the category of experiments, as the experimenters exercise absolute control over the subject and are in a position to introduce stimulus in the varying degrees while keeping other variables stable. In its widest sense, it may mean such manipulation as social legislation which leaves all other causative factors totally uncontrolled. As F. Stuart Chaplin, a pioneer and an authority on experimental type research has aptly said: "while it is a true statement when it means that human beings are not to be physically manipulated in an autocratic or in an arbitrary manner, this does not prevent observation of human relations under conditions of control".

## VARIOUS TYPES OF EXPERIMENTS

E. Greenwood in his work **Research Methods in behavioral Sciences** projects five types of experiments based on the type of setting, nature and extent of control exercised, and the technique adopted for manipulating the variables. These five types of experiments are:

1. **Trial and Error Experiment:** In this method, the experimenter does not have a structured plan of study. He formulates a hypothesis and tries to test it on actual social conditions. As his study is not pre-planned, it is likely to be modified in the light of experiences gained. This method lacks scientific methodology and there is much probability of labour and time going waste.
2. **Controlled Observational Study:** In this method, a stimulus is provided to the subject and changes are observed to find out the causal effect of the stimulus. Experiments in other words, deal with cause and effect. In an ideally designed experiment, there is a direct relationship between the independent and dependent variables. The experimenter concludes with an understanding not only of what happened but why it happened. This observation of the phenomenon under controlled conditions takes this method nearer to laboratory type experiment.

3. **Natural Experiment:** It is also called field experiment because the experimenter conducts his experiments in the natural setting. But it is different from the field-study or field observation because independent variables are manipulated in it, whereas in case of field-study, the researcher simply observes the phenomena without any kind of manipulation.

Experimenters used this method for studying a wide variety of techniques and methods such as participative management, advertising techniques, training methods and effects of pressure groups and the like.

4. **Ex-Post-Facto Technique:** This technique is also called as Field Studies and is helpful in studying the varying influence of two identical factors. If the experimenter wants to study the influence of education being imparted in convent schools and general schools upon future adjustability in life of girls, he shall take an equal number of girl students in each group keeping in view the matching variables like age, economic status, intelligence, etc. After, say a decade, he shall again try to locate those girls and see how they are placed in life. The difference between the two could be attributed to the nature of education received.

Since the ex-post facto technique moves from present to future, often it becomes difficult to locate the subjects of study after the lapse of sufficient time. Stuart Chaplin, therefore, conducted his experiment, backward from present to past while studying the relationship between scouting and delinquency. He took samples from older boys—one consisting of delinquents and the other from non-delinquents. His purpose was to find out how many among the two groups had learnt scouting; Is there any co-relation between scouting and delinquency?

In ex-post-facto design the experimenter moves from cause to the effect, while in Chaplin's technique, he moves from effect to the cause. Chaplin method has been subjected to

severe criticism. It is dubbed as 'imaginary experiment'. When it is required to find out which of the certain results is most likely, the researcher can say it with some certainty, but when the effect is known to us, it is very difficult to say as to which of the various causes has caused it.

5. **Laboratory experiment** : Leon Festinger in his work *Research Method in Behavioural Sciences* defines a laboratory experiment "as one in which the investigator creates a situation with the exact conditions he wants to have and in which he controls some and manipulates other variables". The experimenter operates under the basic assumption that the research situation he wishes to evaluate has never existed and does not now exist. He, therefore, must create it to be able to study it. If experimentation is defined as involving some new and untried element, or condition, then, experimentation involves the evaluation of the effects of this condition or element. At the cost of repetition, the condition or element being evaluated is referred to as the independent variable in the experiment, while the criteria by which it is to be evaluated are referred to as the dependent variables.

The laboratory experiment is different from field experiment in the sense that the latter is studied in the natural setting with control over independent variables, whereas in the former the makes an experiment with the existing classes in the school itself, it would be a field experiment, but if he develops a class room and experiments upon it, it will be a laboratory experiment. In short, in a laboratory experiment the phenomenon is artificially created and subjected to greater control and manipulation. Control is the essential ingredient of experimental method. Its main purpose in an experiment is to arrange a situation in which the effect of variables can be measured.

## PLANNING AN EXPERIMENT

Pre-planning the experiment is of special importance. It would consist of the following steps:

## **1. Selecting a problem**

Every type of research cannot be carried through the experimental method. One of the major conditions for this purpose is the capacity to manipulate independent variable whose effect it is decided to study. Study of motivations on work-efficiency can be cited as an example. Then, the research problem must be stated in clear, precise and conceptual terms. Initial statement of the hypothesis must be made in precise and conceptual terms." The variables or the conditions that affect the phenomenon must be conceptualised. The independent variable should also be decided and the plan for its gradual manipulation be chalked out. In brief, selecting the problem involves the following steps:

1. Statement of the research problem.
2. Determination that the experimental approach is appropriate.
3. Specification of the independent variable.
4. Specification of the levels of the independent variable.
5. Specification of full range of potential dependent variables.
6. Initial statement of hypothesis.

### **Example of Hypothesis:**

Level of over-all job satisfaction under the democratic organizational structure will be high as compared to the over all job satisfaction under the autocratic organizational structure.

## **2. Selecting the Setting**

Setting refers to the back ground in which the phenomenon is to be studied. In case of an experiment in a laboratory, the setting has to be artificially created by the experimenter. In a field experiment, the researcher has to find out the natural setting where the experiment can be made. The following features may be helpful in good setting.

First, the situation must contain the phenomenon to be studied. For example, if we want to determine the influence of democratic and autocratic organisational structures on employees morale and job satisfaction, the experimenter will have to select representative

organisations. Secondly, the phenomenon must occur in sufficiently pure form to enable the researcher to execute his project. Thirdly, the researcher must have the power to manipulate the independent variable. The researcher has limitations in this regard; (a) The phenomenon may be such as does not permit manipulation. (b) The owner of the organisation may not permit manipulation. (c) The researcher must have power to control other factors which might confound the experiment (d) The researcher must have the freedom of access to the data and also freedom to publish the findings.

### **3. Pilot Study**

The test that is administered to the subjects before the independent variable is applied is called pre-test or pilot study. In case of interview schedule, a pre-test is undertaken to be sure that the questions will be as meaningful to the average respondent as to the investigator. It also enables the investigator vis-a-vis realities and may pose problems which the investigator had not even thought of earlier. He will be able to know the various causative factors involved in the nature and working of the organisation chosen for research, the extent of cooperation or resistance he is likely to face.

## **EXPERIMENTAL DESIGN**

An experimental design to the researcher is what a blueprint is to an architect. It enables him to test hypothesis by reaching valid conclusions about relationship between independent and dependent variables. Selection of a particular design is based upon the purposes of the experiment, the type of variables to be manipulated, and the conditions of limiting factors under which it is conducted. The design deals with such practical problems as how subjects are to be assigned to experimental and control groups, the way variables are to be manipulated and controlled, the way extraneous variables are to be controlled, how observations are to be made, and the type of statistical analysis to be employed in interpreting data relationships.

## **EXPERIMENTAL VALIDITY**

To make a significant contribution to the development of

knowledge, an experiment must be valid. D.T. Campbell and J. C. Stanley discusses two types of experimental validity: Internal validity and external validity. T.D. Cook and D.J. Campbell added two other types: statistical validity and construct validity. For our purposes it is better to confine ourselves to the meaning of internal and external validity.

**Internal Validity:** An experiment has internal validity to the extent that the factors that have been manipulated (independent variables) actually have a genuine effect on the observed consequences (dependent variables) in the experimental setting.

**External Validity:** The researcher would achieve little of practical value if these observed variable relationships were valid only in the experimental setting and only for those participating. External validity is the extent to which the variable relationships can be generalised to other settings, other treatment variables, other measurement variables, and other populations.

Control of variables is the basic ingredient of experimental method. It refers to the extent to which different factors in an experiment are accounted for. We have stated that for a successful experiment, it is necessary that the major independent variable should be varied gradually while all other causative factors should remain constant. But in a field experiment particularly dealing with human behaviour, control over causative factors may not be possible. The researcher may vary the quantity of stimulus, but he may not be able to control all other factors from creating confounding effects through variations in them. The following methods are useful in avoiding this limitation.

#### A. *Experimental and Control Groups*

The experimental research involves the comparison of the effects of a particular treatment with that of a different treatment or no treatment. Reference is usually made to an experimental group and to a control group. The group which is subjected to experiment is called experimental group, whereas the other is called the control group. In other words, the experimental group is exposed to the

Randomization is not a haphazard or arbitrary method of assigning objects of a universe to subsets; it is a scientific method of eliminating systematic bias. The principle is based upon the assumption that through random assignment, differences between groups result only from the operation of probability or chance. These differences are known as sampling error or error variance and their magnitude can be established by the researcher.

The major limitation of this method is that perfect randomization is not possible in a small sample and experimentation with a fairly large group would create other complicated problems of its own. Solomon maintains that more than one control group should be selected to eliminate the chances of variation in the group composition during the experimental period. This will also help to verify the results. If two or more groups remain similar after the period of experiment, it will be evident that confounding factors have produced varied influence upon different groups. In fact, randomization provides the most effective method of eliminating systematic bias.

#### ***A.2 Matching Cases***

If the randomization is not possible (e.g., there are few subjects), selecting pairs or sets of individuals with identical or nearly identical characteristics and assigning one of them to the experimental group and the other to the control group, provides another method of control. But the major limitation of this method is that some individuals will be excluded from the experiment if a matching subject is not available. Matching is not considered as satisfactory unless the members of the pairs or sets are randomly assigned to the pair groups, a method known as matched randomization.

#### ***A.3 Balancing Cases or Group Matching***

Balancing cases consists of assigning subjects to experimental and control groups in such a way that the means and the variances of the groups are as nearly equal as possible. But as identical balancing of groups is impossible, the researcher must decide how much departure can be tolerated without loss of satisfactory control. Both these methods of control matching cases and group matching present a

similar limitation; namely, the limitation of equaling groups on the basis of more than one characteristic or variable.

There are certain other methods for controlling variables.

Variables that are not of direct interest to the researcher may be controlled by eliminating them altogether.

#### **A.4 Manipulation**

This is another distinguishing method. It refers to a deliberate operation of conditions by the researcher. In contrast to descriptive research, in which the researcher simply observes conditions as they occur naturally, the researcher in the experimental research actually sets the stage for the occurrence of the factor whose performance is to be studied under conditions in which all other factors which might complicate the observation are controller or eliminated. In the process of manipulation, a pre-determined set of varied conditions referred to as the independent variable, the experimental variable, or the treatment variable, are imposed on the subjects selected for the experiment. Sex, methods of teaching, attitudes, socio-economic status, enthusiasm or teacher's competence, academic ability, personality characteristics, types of motivation, etc. are some examples of research in teaching.

#### **A.5 Observation**

This is still another important method of controlling variable. In the experimental research, the researcher studies the effects of the independent variable on a dependent variable. As already stated, the dependent variable is scores on a test or observations with respect to some characteristics of the behaviour of the subjects used in the experiment.

Despite the use of above methods for controlling variables, some discrepancies invariably remain and influence the result of the experiment. Such discrepancies can be taken care of by the replication of the study.

#### **A.6 Replication**

This is a method of conducting a number of subexperiments

within the framework of an overall experiment design. The researcher, instead of comparing a single control case with a single experimental case, makes a multiple comparisons of a number of cases of the control group, all within the same experimental framework. Repetitive surveys have been more popular than the repetitive experiments. But if laboratory experiments are repeated under similar conditions they can help in confirming the finding of earlier experiment.

## MERITS

Experimental method has been universally hailed as the most rigorous and the most scientific. It has the merit of scientific vigour and mathematical logic as the research is based on a well founded model.

It permits the determination of 'cause and effect' relationship more precisely and clearly in comparison to other methods. It is the best method for testing a hypothesis. Testing of a hypothesis requires a study of relationship under varied conditions which is possible only under this method. The precision and rigorousness of this method makes it more suitable for this purpose.

The method is criticized on the plea that it does not advance knowledge but merely helps in consolidating the facts that have already been existing. This criticism is not tenable because even if this method does not help in theoretical advancement of knowledge, it does provide an exploratory process of testing the existing theories or in identifying the assumptions for improvement or development of an alternative theory. Moreover, consolidating the existing knowledge as a piece of meaningful research is as important as its development.

## LIMITATIONS

This method has its own limitations as given under :

The main argument against this method is whether the results obtained through this method are applicable in the practical situations or whether they are purely hypothetical. The unit of study in this method is so short and the approach is so segmental that chances of its application are rather few. The degree of manipulation of

independent variable pre-supposes that the unit of study must be small and the experiment must be confined to a short period. In those cases where time-span is long, cases may be lost through moving or at least modified. The more highly developed sciences are not time-bound. Difficulty in practical application is also caused through the segmental approach of this method. Human behaviour is not governed by many social causes.

There are many difficulties connected with the selection of setting, in eliciting cooperation from the respondents, and controlling the variables created by complexity of social phenomenon -its dynamic nature and independence of human behaviour. There are also practical difficulties concerning random sampling, loss of cases and shrinkage in numbers in each group during the period of observation and limitations connected with application of probability tests as a means of estimating the significance of the results.

Commenting on Experimental Research, F.N. Kerlinger writes, that social scientific research can be divided into four major categories laboratory experiments, field experiments, field studies, and survey research. he proceeds to highlight the strengths and weaknesses of each category.

### **1. Laboratory experiments**

Kerlinger writes that this experiments have three related purposes. First, they attempt to discover relations under "pure" and 'uncontaminated' conditions. The experimenter asks: is X related to Y? How is it related to Y? How strong is the relation? He seeks to reduce a discovered relation to functional form. A second purpose is mentioned in conjunction with the first purpose; the testing of predictions derived from theory, primarily, and other research, secondarily. A third purpose of laboratory experiments, according to Kerlinger, is to refine theories and hypothesis, to formulate tested hypothesis, and, perhaps most important, to help build theoretical systems.

The aim of laboratory experiments, then, is to test hypothesis derived from theory, to study the precise inter-relations of variables

and their operations and to control variance under research condition that are uncontaminated by the operation of extraneous variables. As such, the laboratory experiment is one of man's greatest achievements.

Highlighting the strengths of laboratory experiments, Kerlinger writes that the laboratory experiment has the inherent virtue possibly of relatively complete control. In addition to situation, control, laboratory experiments can ordinarily, use random assignments and can manipulate one or more independent variables. Moreover the experimenter in most cases can achieve a high degree of specificity in the operational definitions of field situations, success as many of those associated with the measurement of value attitudes, aptitudes, and personality traits. Closely allied to operation strength is the precision and responsibility of laboratory experiment. Precise means accurate, definite, and unambiguous. Precise measurements are made with precision instruments. In variant terms, the more precise an experiment procedure is, the less error variance. Precise laboratory results are achieved mainly through controlled manipulation.

The greatest weakness of the laboratory experiment is probably the lack of strength of independent variables. Since laboratory situations after all are the situations that are created for special purposes, it can be said that the effects of experimental manipulations are usually weak. Another weakness is the corollary of the first: the artificiality of the experimental research situation. The criticism of artificiality of the experimental research situation. The criticism or artificiality comes from individuals lacking an understanding of the purposes of laboratory experiments. Moreover, the temptation to interpret the results of laboratory experiments incorrection is great. Furthermore, although laboratory experiments have relatively high internal validity, then, they lack external validity.

Although weaknesses exist, they are the waeknesses only in a sense that is really irrelevant. Conceding the lack of representativeness (external validity) the laboratory experiment still has the fundamental pre-requisite of any research: internal validity.

## 2. Field Experiment or Natural Experiment

This is a research study in a realistic situation in which one or more independent variables are manipulated by the experimenter under as carefully controlled conditions as the situation permit. The contrast between the laboratory experiment and the field experiment is not sharp the differences are mostly matters of degree. Sometimes it is difficult to label a particular study "laboratory experiment" or "field experiment". Where the laboratory experiment has a maximum of control, most field studies must operate with less control, a factor that is often a severe handicap to the experiment.

Field experiments have values that especially recommend them to social psychologists, sociologists, and educators because they are admirably suited to many of the social and educational problems. The control of experimental field situation, however, is rarely as tight as the laboratory experimental situation. The necessity of controlling extraneous independent variables is particularly urgent in field experiments. Another virtue of field experiment is their appropriateness for studying complex social influences, processes, and changes in lifelike settings. Moreover, field experiments are well suited both to the testing of theory and to the solution of practical problems. Still another characteristic of field experiments is that they are suited to testing broad hypothesis. Finally, flexibility and applicability to a wide variety of problems are important characteristics of field experiments.

The main weakness of field experiments are practical. Manipulation of independent variables and randomization are perhaps the two most important problems. Another weakness inherent in field experimental situations is lack of precision as is possible in the laboratory experiment. In realistic situations, there is always a great deal of systematic and random noise.

Field Studies are ex-post-facto scientific inquiries aimed at discovering the relations and educational variables in real social structures. Such studies are strong in realism, significance, strength of variables, theory orientation, and heuristic quality.

Despite these strengths, the field study is a scientific weak cousin of laboratory and field experiments. Its most serious weakness, of course, is its ex-post-facto character. Thus statements of relations are weaker than they are in experimental research. To complicate matters further, the field situation almost always has a plethora of variables and variance. Another methodological weakness is the lack of precision in the measurement of field variables. In field studies, the problem of precise is more acute, naturally, than in field experiments. Other weaknesses of field studies are practical problems; feasibility, cost, sampling, and time. These difficulties are really potential weakness-none of them need be a real weakness. In designing research, it is important not to underestimate the large amounts of time, energy, and skill necessary for the successful completion of most field studies. The field researcher needs to be a salesman, administrator, and entrepreneur, as well as investigator.

Nevertheless, one must conclude by saying that the use of experimental method in social sciences is limited because of the difficulties posed by social phenomenon and researcher's lack of knowledge of it rather than due to any inherent defect in the systems. As our knowledge of the social phenomena increases, our researchers will be in a better position to make increasing use of this method. In the meantime, we may agree with the appraisal of this method by John French made in his study, **Experiments, In Field Settings** that "Through careful measurement, better theorizing, the use of control groups, and other aspects of improved experimental design the practical problems of social action can be solved with greater certainty, with greater accuracy, sometimes, with greater efficiency than through common sense trial and error method".

## **DIFFERENT TYPES OF RESEARCH : INTER-DISCIPLINARY RESEARCH**

Inter-disciplinary approach is a technique of research in which the tools of different sciences are used to find an explanation to the phenomenon under study. It is thus a cooperative and coordinated research in which the experts of different disciplines pool their talents together for the purpose of finding explanation to any problem. As Pauline V. Young has succinctly put; Ideally speaking, cooperative research implies a group of specialists who can coordinate their work in a consistent attempt to keep within common scientific interests rather than to pursue lines of specialist relevance.

The demand for inter-disciplinary research is increasing. It is because the universe of subjects at bottom is a continuum. It means that whatever the atomizing intellect might do, no subject can be developed without its calling for some development in other subjects sooner or later. Further, reasons accounting for the mounting need for interdisciplinary research are not far to seek.

### **NEED OF INTER-DISCIPLINARY RESEARCH**

First, the mode of research has changed. The modern scientists owe great responsibility to the society. It is to science that the emerging nations look for their rapid economic, social, cultural, religious, technological and political development; and the developed countries

for their economic stability. Conversely the progress, welfare, and security of the nations depend critically on rapid, planned and sustained growth in the quality and extent of education and research in science and technology.

Till the middle of the nineteenth century, scientists were pursuing research mostly to satisfy their inner urge and it was spontaneous. But now the pressure is on the scientists to bring their labour to bear upon the needs of the mankind. It has relegated to the background that individuality of the scientists and team research has taken its place. Let us consider the rubber industry in this connection. How many are the scientists involved! In the cultivation of the rubber tree, the botanists and the mycologists are perhaps of great importance. The subsequent stages of manufacture demand the specialist knowledge of chemist, the physicist, and the engineer. In addition, the business side calls for expert statisticians, accountants, production engineers, personnel officers, public relations men, and management specialists. Thus uni-disciplinary research is yielding place to multi-disciplinary research.

Secondly, upto the close of the nineteenth century, the research in various disciplines was mostly hierarchical. But now we have ushered in an era of inter-disciplinary research or what Einstein, calls the "Unity of Science". The case of Space Technology provides an illustration; It is the outcome of continued researches in Astrophysics, Nuclear Physics, Astronomy, Geology, Microbiology, Medicines, Engineering, and telemetry etc. This inter-disciplinary research has given birth to new disciplines like Space Medicine, Space Biology and Space Geology, etc.

Similarly, with the advances in Physics, Chemistry and Technology, the field of Biological Sciences has greatly benefited. The theories propounded by great scientists like Hargobind Khorana and Anand Chakrabarty have caused the emergence of new subjects like Genetic Engineering. Application of Chemistry and Physics has resulted into Bio-chemistry and Bio-physics respectively. Microbiology and Molecular Biology are other instances. On one hand, the inventions are affecting the structure of knowledge and

on the other hand, they are approaching towards inter-disciplinary approach. The most fruitful results in research are achieved not only through an integration of scientific techniques and method but also through a unified approach of the various scientific disciplines.

Thirdly, in social sciences, too, inter-disciplinary approach is coming into vogue. Since man lives in a world of economic, industrial, political, psychological forces and social attitudes and values, it is self-evident that his responses to, and roles in these should be studied. As a matter of fact, the Indian Council of Social Science Research was established in 1969, to promote and integrate the social sciences to each other and to the related useful arts of industry, government, and public welfare, to guard against overspecialisation and isolated efforts which might result in a distorted vision.

We are living in age of specialization, there is craze for knowing more and more of a social problem. The various facets of a problem can only be studied when there is interdisciplinary research and people belonging to different disciplines study, the problem from their own view-point and produce the results.

Fourthly, of course, specialisation in one particular discipline is essential for healthy growth and progress of science. But excessive adoration and submissiveness to one's own discipline and specialisation in one particular branch of social science, will only give rise to the narrow-minded, dogmatic and egoistic approach. Charles E. Kellogg aptly observed; "even the scientist is laymen in most fields. If he simply uses the special techniques of his own field without undertaking the deeper meaning of the scientific method as applied generally, he may be as blindly dogmatic or as bad a dupe as anyone in other fields." Inter-disciplinary approach is essential to avoid the evils of over-specialisation.

Fifthly, one of the most significant contributions in cooperative research, according to Pauline V. Young, is that it facilitates the study and analysis of the complex web of social-psychological economic forces intricately inter-woven in modern life. Although different disciplines are separated from each other, yet they overlap on many points. This is because the central focal point for all the

sciences is man. This overlapping of the sphere of operation makes different disciplines interdependent and thereby creates need for inter-disciplinary approach.

Sixthly, cooperative research makes the whole study objective. An individual researcher pursuing his own line of thought may be highly biased and dogmatic in his approach. He may deliberately or inadvertently try to project his own theories into the phenomena to find a possible explanation rather than making an objective study of the same. Free and open constructive criticism of informed colleagues is essential to arrive at sound conclusions. Inter-disciplinary approach brings a person out of a particular discipline and take a more comprehensive view of the problem. He takes into consideration others' view-point on the problem and this helps him in reaching objectivity. Without scrutiny and earnest criticism by others interested in the same problem, a scientist cannot presume to speak with the authority.

### **PRE-REQUISITES OF INTER-DISCIPLINARY RESEARCH**

The first pre-requisite of the inter-disciplinary research is the formulation of a definite and concrete problem of study dealing with more than one subjects. If the problem under study is uni-disciplinary, then, multi-disciplinary approach will be meaningless. Secondly, the problem selected for research should describe clearly the objectives and scopes of research. This will help in deciding the type of experts that would be needed to assist in the execution of the project. The third pre-requisite would be to select a team of trained and competent experts who can undertake the work in each discipline connected with the study. They should be persons of vision and liberal thinking and should not be unduly dogmatic, narrow-minded or egoistic. They should be willing to cooperate with the plan on equal terms and are prepared to work with some common objectives. Pauline V. Young, has aptly said; "Cooperative work presupposes disciplined, united efforts of persons who will not be distracted by self-interests, logical arguments, or jealousies." All should believe that their findings will benefit the society; that they are working for the welfare and advancement of society.

## PROCESS OF INTER-DISCIPLINARY RESEARCH

Inter-disciplinary research can be completed in three different ways:

1. The social scientist may take upon himself the responsibility of undertaking the whole job. He picks a team of field investigators belonging to different disciplines but competent enough to use different tools of research, assigns them specific responsibilities, directs them and seeks the help of people belonging to other disciplines and their specialists as and when needed. Or he may carry on the whole research work himself consulting occasionally the different experts.
2. The collaborators who are interested in the problem have no coordination in the beginning. They carry on their studies independently. In the end their findings may be coordinated and integrated to prepare a combined research report.
3. An inter-disciplinary problem on which research is to be conducted is picked up by the researchers. They finalise the design and each one then picks up the aspect or dimension of the problem which he can confidently study. The researchers associated with the problem, however, meet from time to time and review the design and made changes, if needed.

## MERITS OF INTER-DISCIPLINARY RESEARCH

Inter-disciplinary approach has a number of commendable points to its credit.

First, the problem can be studied as a unified whole in all its aspects; it helps to avoid the defects of the segmental approach. Moreover, when each aspect is being studied by an expert in the field, he would do maximum justice with the subject. Therefore, the problem will be studied thoroughly and from all dimensions.

The second advantage is that the study becomes, well balanced; each discipline is represented in it.

Thirdly, inter-disciplinary approach removes the dangers of overspecialisation, and the chances of bias caused by undue adherence

and dogmatic clinging to a particular discipline are reduced to the minimum.

Fourthly, it helps in securing objectivity. The researcher has to proceed with wide vision, ready to learn from other disciplines. This makes him free from any biased conclusion.

Fifthly, it provides a scope for the study of comparative efficacy of the different disciplines. When the same problem is studied by the experts of different disciplines we can know the degree of similarity and conflict in the approach and the results obtained. This may lead to unification of different disciplines.

Sixthly in an inter-disciplinary research, advantages of different tools of research are available.

Finally, this approach of research develops a feeling of understanding and a spirit of cooperation among the experts of different disciplines. They understand restraints, and constraints of each discipline and try to understand each others' problems and view-points, so that they can work in close harmony and cooperation with each other. It helps to remove the dogma and conservatism from them and make them more broad-minded.

## **LIMITATIONS OF INTER-DISCIPLINARY RESEARCH**

The following are the limitations or demerits of this type of research.

Inter-disciplinary research has not gone without challenges. It is very difficult, if not impossible, to bring together the experts of different disciplines to work on a common problem. Every discipline has a different approach, to the problem and a methodology suited for the purpose. The study is likely to become "haphazard" and "unfocused on any aspect".

Secondly, it is usually difficult to find out a subject of study for which interested researchers are available. Sometimes, subject of study is selected simply because the people are available and in this way no good research work is possible.

Thirdly, even if the researchers are available, many times,

good research problems are not solved satisfactorily because the researchers do not work in close cooperation with each other or just pull them in the opposite direction.

Fourthly, competent and trained field investigators are required when a comprehensive research problem is taken up for study. But in actual practice it is very hard to find trained and competent field investigators, who can provide proper feed back to the principal investigator or project directors for completing their work.

Fifthly, in this approach of research, the chances of delay are numerous because too many persons, working at different levels are involved.

Sixthly, in inter-disciplinary research huge economic and human resources are needed to study the various aspects of the problem. But in every society it is very difficult to raise the resources for one study alone. This problem is more acute in case of developing countries which have neither economic resources nor research expertise.

Notwithstanding its limitations, the utility of inter-disciplinary research in modern society cannot be under-estimated, because it is now increasingly felt that unidisciplinary research is more or less not possible keeping in view the proliferation of subjects. "In fact, as we see it today more clearly, the social sciences to be truly scientific must stand or fall-together. The old individualism is in retreat, and social scientists are becoming socialized."

Concluding in the words of A. Neelamlghan, it can be said that as a method of discovery, research sets no boundaries to the subject of investigation. In the process of the development of a subject, its interaction with the environment becomes more and more extensive as well as capability to utilise and assimilate the knowledge generated in other disciplines. This leads to the mutual enrichment of the subject, the development of new specialisations, recognition of new problems, further research, and so on. The development of research through such cross-fertilisation of ideas is really commendable.

# 5

## SCIENTIFIC METHODOLOGY

### WHAT IS SCIENCE?

Science is popularly defined as a systematic study of knowledge. The words 'systematic' and 'knowledge' need further elaboration. Knowledge refers to the goal of science, while systematic signifies the method used in achieving that goal. In fact, the main objective aim of any type of study-whether scientific or otherwise is the acquisition of knowledge, to know the truth and reality behind a phenomenon. Thus, the word 'systematic' distinguishes science from other branches of knowledge. The different interpretations of the word 'systematic' has caused a lot of confusion in deciding whether any particular branch of knowledge called science or not ?

Before defining, what is scientific method and how it works, it is better to put the word 'science' in the right perspective.

The practitioners of the Natural Sciences had monopolized the term 'Science' for a long time. According to them a science should be restricted to:

1. Subjects dealing with the physical world or its phenomena.  
or
2. Subjects dealing with the world of living bodies and their phenomena.

Joseph A. Schumpeter broadens the restricted definition of

science and he writes: "a science is any kind of knowledge that has been the object of conscious efforts to improve it. Such efforts cultivate habit of mind-methods or 'techniques'-and a command of facts unearthed by these techniques which are beyond the limit of mental habits and the factual knowledge of everyday life. Hence we may also adopt the practically equivalent definition; a science is any field of knowledge that has developed specialized techniques of fact-finding and of interpretation or inference (analysis). Finally, if we wish to emphasize sociological aspects, we may formulate still another definition, which is also practically equivalent to the other two; a science is a study of any field of knowledge in which there are people, research workers or scientists or scholars, who are engaged in the task of improving upon the existing stock of facts and method and who, in the process of doing so, acquire a command of both that differentiates them from the 'layman' and eventually also from the mere 'practitioner'.

Speaking in the same terms, Karl Pearson also writes that "The material of Science is unlimited - every group of natural phenomena, every phase of social life, every stage of past and present development is material (for science)... The material of science is co-extensive with the whole of life". Pearson, in fact, highlights that the unity of all sciences consists alone in its method, not in its materials. In his own words; "There is no shortcut to truth.. no way to gain knowledge of the universe except through the gateway of scientific method". The "Scientific Method," according to Pearson, "is one and the same in all branches and that method is the method of all logically trained minds. The unity of all sciences consists alone in its methods not its material; the man who classifies facts of any kind whatever, who sees their mutual relation and describes their sequences, is applying the scientific method and is a man of science." It is not the facts themselves which make science; it is rather the method by which they are dealt with that make a science. Science is independent of any particular subject-matter or order of facts. "It takes knowable Universe for its subject. It deals with physical as well as psychical processes, with man as well as with

nature. It has to do with everything to which its methods can be applied. What makes a Science is not, of course, the nature of things with which it is concerned but the method by which it deals with things. Study of a skylark is not necessarily Zoological". Stuart Chase in his book the "Proper Study of Mankind", too, supports this view when he writes, "Science goes with the method and not with the subject matter". All that the term science means is a field which has been studied according to certain principles, i.e., according to scientific method, says Lundberg. S.R. Ranganathan defines science as the name given to any domain in the Universe of Knowledge, whose development is characterised by a method called "Scientific Method".

A. Wolfe in "Essentials of Scientific Method" lays bare the characteristics of Scientific Method. To him; "Any mode of investigation by which science has been built up and is being developed is entitled to be called a science: (a) critical discrimination; (b) generality and system, and (c) empirical verification. Any method of study by which the above objectives can be achieved may be termed as scientific method.

The foregoing analysis of the terms science and scientific method may be summed up; Science is an objective, logical and systematic method of analysis of phenomenon devised to permit the accumulation of reliable knowledge. It is a systematized form of analysis, not any particular body of knowledge. In the words of Will Durant, "Science is the captured territory (in the pursuit of truth). As soon as a field of enquiry yields knowledge susceptible of exact formulation, it is called Science." The scientific method refers to a procedure or a mode of investigation by which scientific and systematic knowledge is acquired. In fact, any study can claim to be called scientific wherein the data are subjected to logical analysis regardless of the fact whether the results are obtained by experiment, statistics or common sense. David J. Luck aptly writes that any research to be called scientific must satisfy the following requirements:

- \* that it is an orderly investigation of a defined problem;
- \* that approximate scientific method be used;

- \* that adequate and representative evidence be gathered;
- \* that logical reasoning, uncoloured by bias, be employed in drawing conclusions on the basis of evidence;
- \* that the researcher be able to demonstrate or prove the validity of reasonableness of his/her conditions/thesis; and
- \* that the cumulative research in a given area yield principles or laws that may be applied with confidence under similar conditions in the future.

Francis Bacon (1561-1626) planted the seeds of the scientific method when he attacked the **deductive method**. For reaching conclusions on the basis of observed facts, by deductive method we always arrive at a conclusion from general to specific. Man devised inductive reasoning to complement deductive reasoning as a means of searching knowledge. In inductive reasoning, an investigator collects evidence that will enable him to establish a generalisation as being probably true. He initiates his inquiry by observing particular instances or concrete facts, and from his examination establishes general conclusion about the whole class in which these particular instances belong. Each method has its merits and demerits.

To construct a more practical method of attaining reliable knowledge, men such as Newton, Galileo, and their successors eventually combined the inductive and deductive thought processes. This synthesis of reason and observation produced the modern scientific method of research.

### **WHAT IS SCIENTIFIC METHOD?**

George L. Lundberg says that: "Scientific method consists of systematic observation, classification and interpretation of data. The main difference between our day to day generalisation and the conclusion usually recognized as scientific method lies in the degree of formality, rigorousness, verifiability and general validity of the latter."

Karl Pearson gives an exhaustive definition. To him the "Scientific Method is marked by the following features:

1. Careful and accurate classification of facts and observation of their correlation and sequence;
2. The discovery of scientific laws by aid of creative imagination; and
3. Self-criticism and the final touch stone of equal validity for all normally constituted minds.

L.L. Bernard, states in "Field and Methods of Sociology", Scientific Method consists of six process that take place within it. These are testing, verification, definition, classification, organisation and orientation, which includes predication and application."

The purpose of research is to discover answers to questions through the application of scientific procedures. These procedures have been developed in order to increase the likelihood that the information gathered will be relevant to the questions asked and will be reliable to unbiased. Scientific Method involves certain stages which invariably leads to systematic, controlled, empirical, and critical investigation of hypothetical propositions about the presumed relations among natural phenomena. However, two points need to be emphasised. First, when we say that scientific research is systematic and controlled, we mean in effect that scientific investigators can have critical confidence in research outcome. His means that research situation in tightly disciplined. Secondly, scientific investigation is empirical. If the scientist believes something to be so, he must somehow other put his belief to a test outside himself. Subjective belief, in other words, must be checked against objective reality. As such scientific research means a "systematic, frequentive and intensive study of collected data."

## **COMPONENTS/STAGES INVOLVED IN SCIENTIFIC METHOD/INQUIRY**

### **1. A Felt Difficulty**

Man encounters some obstacle or problem that puzzles him, may be, that (a) he lacks the means to get to the desired; or (b) he has difficulty in indemnifying the character of an object; or (c) he

cannot explain an unexpected problems as India is a rich country inhabited by poor? Why the First War of Indian is a rich country inhabited by poor? Why the First War of Indian Independence (1957) Failed? Why an apple falls down and does not float in the air?

## 2. Definition of the Problem

A scientific inquiry starts with the identification of a problem that is in need of solution. The problem identified must be defined in such a manner that observation or experimentation in the natural world can provide a solution. According to F.L. Whitney; "To define a problem means to put a fence around it, to separate it by careful distinctions from like questions found in related solutions of need. "The researcher must know his problem precisely and narrowly it down to workable size before he starts his works on it. As W.S. Monroe and M.D. Engelhart aptly remark; "To define a problem means to specify it in detail and with precision. Each question and subordinate question to be answered is to be specified. The limits of the investigation must be determined. Frequently, it is necessary to review previous studies in order to determine just what is to be done. Sometimes it is necessary to formulate the point of view or educational theory on which the investigation is to be based. If certain assumptions are made, they must be explicitly noted".

Tyrus Hillway in his book, "Introduction to Research" has suggested the following rules to the researchers for deciding a final definition of a problem.

1. Be sure that the topic chosen is neither too vague nor too broad in scope.
2. To make the problem clearer and more understandable; state it as a question which requires a definite answer.
3. Carefully state limits of the problem, eliminating all aspects and factors which will not be considered in the study.
4. Define any special terms that must be used in the statement of the problem.

In brief, at the stage of definition, the researcher is required

to describe the background of the study, its theoretical basis and underlying assumptions, and state the problem in concrete, specific, and workable question; it must clarify exactly what is to be determined or solved. The more insight the researcher has into the problem the simple will be his hypothesis about it.

### **3. Formulation of Hypothesis**

The next most important step in the research process is the formulation of the hypothesis. A hypothesis is a proposition, condition, or principle which is assumed, perhaps without belief in order to draw out its logical consequences and by this method to test its accord with facts which are known or may be determined. Hypothesis, according to Goode and Hatt, "is a proposition which can be put to test to determine its validity. It may seem contrary to, or in accord with, common sense. "Speaking in the same veins, George A. Lundberg writes that "An hypothesis is a tentative generalization, the validity of which remains to be tested. In its most elementary stage, the hypothesis may be any hunch, guess, imaginative idea, which becomes the basis for action or investigation." To F.N. Kerlinger: "An hypothesis is a conjectural statement, of the relation between two or more variables. Hypothesis are always in declarative form and they relate, either generally or specifically, variables to variables." W.H. Werkmeister writes that "The guesses he (the investigator) makes are the hypothesis which either solve the problem or guide him in further investigation".

Hypothesis is in a sense a type of supposition or assumption or immigrating, -a generalisation in the form of a proposition derived on the basis of induction, i.e. reasoning - which may or may not be true. Even a negative result, under certain circumstances, can be the crow of research. But without formulating a hypothesis, scientists can not know what things to observe and what to ? As Lundberg rightly observes; "The only difference between gathering data without a hypothesis and gathering them with one is that in the latter case we deliberately recognize the limitations of our senses and attempt to reduce their fallibility by limiting our field of investigation so as to

permit a greater concentration of attention on the particular aspects which past experience leads us to believe are significant for our purpose".

The hypothesis is a very powerful tool in research process to achieve dependable knowledge; the success or failure of a research work depends upon how best it has been formulated. In fact, it is difficult to conceive modern science in all its rigorous and disciplined fertility without guiding power of hypothesis. Hypothesis originate in the science itself.

The hypothesis helps the researcher to relate theory to observation and observation to theory. Theory is, in fact, an elaborate hypothesis which deals with more types of facts than does the simple hypothesis in actual practice, the relation between the two is so close that it is difficult to separate one from the other. As Goode and Hatt write; "Hypothesis, when tested, are either proved or disproved and in turn constitute further tests of the original theory." In either case, they may be of use to existing theory and may make possible the formulation of still other hypothesis. But such a simple outline, unfortunately, fails to indicate that the formulation of useful hypothesis is one of the most difficult steps in scientific method.

Goode and Hatt have identified three major difficulties in the way to the formulation of useful hypothesis. First among these is the absence of a clear theoretical framework. Second is the lack of ability to utilize that theoretical framework logically. Third is the failure to be acquainted with available research techniques, so as to be able to phrase the hypothesis properly.

### ***3.1 Virtues of Hypothesis***

D. Ary et. al. in their work, 'Introduction to Research in Education' have highlighted the importance of hypothesis as under:

1. Hypothesis facilitate the extension of knowledge in an area. They provide tentative explanations of facts and phenomena, and can be tested and validated. Such explanations, if held valid, lead to generalisations which help significantly in

understanding a problem and thereby extend the existing knowledge in the area to which they pertain.

2. Hypothesis provide the researcher with rational statements, consisting of elements expressed in a logical order of relationships, which seek to describe or to explain conditions or events, that have not yet been confirmed by facts. The hypothesis enable the researcher to relate logically known facts to intelligent guesses about unknown conditions.
3. Hypotheses provide direction to research. They represent specific objectives and thus help the researcher to determine the type of data need to test the proposition. Hypothesis provide a basis for selecting the sample and research procedure to be used in the study. The statistical techniques needed in the analysis of data, and the relationships between the variables to be tested, are also implied by the hypothesis. Moreover, the hypothesis help the researcher to delimit his study in scope.

Finally, hypothesis provide the basis for reporting the conclusions of the study. On the basis of these conclusions, he can make the research report interesting and meaningful to the readers.

### **3.2 Types of hypotheses**

Goode and Hatt have enlisted the following types of hypothesis:

1. Some hypothesis state the existence of empirical uniformities.
2. Some hypothesis are concerned with complex ideal types. These hypothesis aim at testing the existence of logically derived relationships between empirical uniformities.
3. Some hypothesis are concerned with the relations of analytic variables. These hypothesis occur at a level of abstraction beyond that of ideal types. Whereas the hypothesis of empirical uniformities lead to the observation of simple differences and those dealing with ideal types lead to specific coincidences of observation. The study of analytical variables requires the formulation of a relationship between changes in one property and changes in another.

4. The general culture in which a science develops furnishes many of its basic hypothesis.
5. Hypothesis originate in the science itself.
6. Analogies are often a source of useful hypothesis.
7. Hypothesis are also the consequence of personal, idiosyncratic experience.

### 3.3. Characteristics of Usable Hypotheses

Goode and Hatt write that in the privacy of the scientists mind alone, or in social gatherings, in odd moments or in the pressure of business, many philosophy are entertained. Most of them, having appeared, are fortunately left or die alone. A few survive to be tried out. Most are not destined to play any significant role in the growth of science. It is only by the imposition of firm standards that it is possible to win now out the good ideas from the bad.

The learned authors give some criteria for judging the hypothesis as under:

1. The hypothesis must be conceptually clear.
2. Hypothesis should have empirical reference. Not usual hypothesis can embody moral judgements.
3. Hypothesis should be related to available techniques.
4. The hypothesis should be related to a body of theory.
5. The hypothesis should be such that kit can be put to empirical test; and

Finally, the hypothesis should be simple and to the point, The more insight the researcher has into the problem, the simple will be hypothesis bout it.

### 4. Collection of data

After the formulation of the hypothesis, collection of data is of utmost importance for the researcher. If the data are inaccurate and inadequate, the findings of the study are bound to be misleading, howsoever, a scientific research might have been.

Data may be gathered on the problem of within a selected

field by observation method. Observation can fairly be called the classic method of scientific inquiry, a scientific procedure for comprehending things, events, quantities, qualities and relationships.

Three modes of observation may distinguished;

#### **4.1 Visual Observations**

Data are recorded through sense experiences but sense experience is subjective. Science results in unrelinquishable search for objectivity. Science reaches the objectivity in three ways:

- (a) The observation of different persons similarly studied must lead to the same result.
- (b) It must be repeatable by the same persons or others.
- (c) The quality of observed data by one should agree with that of observed by another.

#### **4.1.2 Aid of Scientific Research Instrument**

Observation by human sense must be assisted with suitable instruments like telescope, and photographic plate because these instruments, help in recording facts which would be otherwise unobservable.

#### **4.1.3 Experiment**

In nature, the conditions under which we have to observe a phenomenon may occur in its own time. But in an experiment we can arrange the conditions under which a observation is to be made. Thus experimental method makes possible a more reliable analysis that the observer may repeat the experiment over and over again at will and thus greatly lessen the probable error of his determinants. Few merits of the experimental method are worth mentioning. First, results are more reliable. Second, the whole experiment can be divided to suitable parts or stages to make it a detailed observation. Third, it can be studied or observed under a variety of conditions. Finally, it can be observed by several people at several places.

Be it noted that the observation is only one of the methods. The other methods for collecting data are:

(i) Interview Method; (ii) Schedule Method; (iii) Questionnaire Method; (iv) Case/Life History Method; (v) Experimental Method, etc.

#### **4.2 Analysis of Data**

At this stage the collected data are classified and arranged on the basis of the similarities, variations, activities, processes, causes and results etc. Distinction is also made between essential and superficial characteristics.

#### **5. Testing of Hypothesis**

Testing of a hypothesis or verification of generalisation is done by applying the method of deductive inference. Deductive inference is reasoning from general premises to specific derived instance thereof. The procedure involved may be summarised in as simple hypothesis. If proposition (p) is true, then, proposition (q) is likewise true. If it rains, the streets will be wet (if p then Q). This method of deductive inference proceeds in the following fashion.

- \* the general principle is stated;
- \* a special case is deducted as an instance of general principle; and
- \* inference, conclusion, proof or demonstration results there from.

By this method a hypothesis is tested and accepted in two ways:

1. a hypothesis must tally with the already existing theories;
2. the consequences derived from it must tally with experience and observation.

Consequences from this hypothesis are derived by means of deductive inference. This basis must be said to be logical basis of testing a hypothesis. Thus, when we say that a hypothesis is verified, it means that hypothesis explain the concepts or facts which it set out to explain. In science, if may mean that hypothesis gives the cause of the phenomena explained. Hypothesis are verified in science

by controlled experiments, by tested predictions of results, by repetition of experiments and the gathering of additional data.

The Scientific Method, thus consists of three processes observation, formation of hypothesis and testing of hypothesis. It is one of the most important and the most recognised method for developing knowledge. When a hypothesis is tested and proved, it becomes the theory.

## **6. Reporting**

It means subjecting results to criticism and verification of other competent collaborators.

## **7. Disseminating the results of the research**

Finally, the results of the researchers are disseminated to general public for practical use.

## **SUMMARY OF SCIENTIFIC METHODOLOGY**

An experiment starts with the formulation of a problem simply, clearly, and completely. The researcher may often have only a rather general, diffused, even confused notion of the problem. This is in the nature of the complexity of scientific research. It may even take an investigator years of exploration, thought and research before he can even clearly say what questions he has been seeking answers to. Nevertheless, adequate statement of the research problem is one of the most important part of research. Another part lies in knowing what a problem is and especially what a scientific problem is ? The only requirement that the problem must meet is that it be solvable-the question that it raises must be answerable with the tools available. Nonetheless, there are three criteria of good problems and problem statements. First, the problem should express a relation between two or more variables. Second, the problem should be stated clearly and unambiguously in question form. The third criterion demands that the problem and the problem statement should be such as to imply possibilities of empirical testing.

The researcher formulates a tentative solution to the problem. This tentative solution is called a hypothesis; it may be reasoned

potential solution or only a vague guess, but in either case it is an empirical hypothesis in that it refers to observable phenomena. Hypothesis are statements about the relations between variables; they carry implications for testing the stated relations.

Following the statement of the hypothesis, the researcher tests it to determine whether the hypothesis is (probably) true or (probably) false. If true, it solves the problem. Doubtless, hypothesis are important and indispensable tools of scientific research. Three main reasons can be advanced for this belief. First, they are so to speak, the working instruments of theory. Second, hypothesis can be tested and shown to be probably true or probably false. Third, hypothesis are powerful tools for the advancement of knowledge because they enable man to get outside himself. Though constructed by man, hypothesis exist, can be tested, and can be shown to be probably correct or incorrect apart from man's values and opinions. This is so important that one can venture to say that there would be no science in any complete sense without hypothesis. Hypothesis advances scientific knowledge by helping the investigator to confirm or deconfirm the theory.

To test the hypothesis, we must collect data, for a set of data is our only criterion. Various techniques are available for data collection but experimentation is the most powerful. A test is made of the hypothesis by confronting it with the data. We organize the data through statistical methods and make appropriate inferences to determine whether the data support or refute the hypothesis. Assuming that the hypothesis is supported, an additional step involved in the scientific method is generalization of all things with which the hypothesis is legitimately concerned, in which case we should explicitly state the generality with which we wish to advance the hypothesis. The next step in the Scientific method, closely related to generalization, concerns making predictions on the basis of the hypothesis. By this we mean that a hypothesis may be used to predict certain events in new situations, to new events not studied in original experiment. In making a prediction we may test the hypothesis, a new in the novel

situation, that is we might replicate. Finally we should attempt to explain our findings by means of a more general theory.

## SCIENTIFIC POSTULATES

Researchers working in the natural sciences usually proceed on the basis of some or all of the following assumptions, postulates, or conditions.

1. The principle of causality is the belief that every event has a cause and that the same cause always produces the same effect.
2. The principle of predictive uniformity involves the assumption that a group of events will show the same degree of interconnection or relationship in the future as has been shown in the past or is being shown in the present.
3. The principle of objectivity requires the investigator to be objective with regard to the data before him. The facts must be such that they can be experienced in exactly the same way by all normal persons. The aim is to eliminate all subjective and personal elements.
4. The empiricism lets the investigator assume that his sense impressions are correct and that the test of truth is an appeal to the "experienced facts".

# 6

## BASIC ELEMENTS OF THE SCIENTIFIC AND SOCIAL SCIENCES METHODS

### HYPOTHESIS AND ITS FUNCTION

Facts, as has been shown, are dependent upon a theoretical framework for their meaning. They are also statements of relationships between concepts. A basic requirement in the application of the scientific method, the clear definition of concepts, has just been discussed. The next step, how to ask the questions which lead to new scientific propositions, must now be considered.

How, the theory serves to order and give meaning facts? It has been discussed already. It also pointed out that theory can give direction to the search for facts. A hypothesis states what we are looking for. When facts are assembled, ordered, and seen in a relationship, they constitute a theory. The theory is not speculation, but is built upon fact. Now, the various facts in a theory may be logically analyzed, and relationships other than those stated in the theory can be deduced. At this point there is no knowledge as to whether such deductions are correct. The formulation of the deduction, however, constitutes a hypothesis; if verified it becomes part of a future theoretical construction. It is thus clear that the relation between the hypothesis and theory is very close indeed. One scientist, in this

connection, has stated: "In practice a theory is an elaborate hypothesis which, deals with more types of facts than does the simple hypothesis. . . . The distinction . . . is not clearly defined." While it is true that the two can never be satisfactorily separated, it is useful to think of them as two aspects of the way in which science adds to knowledge. Thus a theory states a logical relationship between facts. From this theory other propositions can be deduced that should be true, if the first relationship holds. These deduced propositions are hypothesis.

A hypothesis looks forward. It is a proposition which can be put to a test to determine its validity. It may seem contrary to, or in accord with, common sense. It may prove to be correct or incorrect. In any event, however, *it leads to an empirical test*. Whatever the outcome, the hypothesis is a question put in such a way that an answer of some kind can be forthcoming. It is an example of the organised skepticism of science, the refusal to accept any statement without empirical verification.

Every worth-while theory, then, permits the formulation of additional hypothesis. These, when tested, are either proved or disproved and in turn constitute further tests of the original theory. In either case they may be if use to existing theory may make possible the formulation of still other hypotheses. Such a simple outline, unfortunately, fails to indicate that the formulation of useful hypotheses is one of the most difficult steps in scientific method.

## PROBLEMS IN FORMULATING THE HYPOTHESIS

As difficult as the process may be, it is necessary for the students to see the fundamental need of a hypothesis to guide sound research. Without it, research is unfocused, a random empirical wandering. The results cannot even be stated as facts with a clear meaning. The hypothesis is the necessary link between theory and the investigation which leads to the discovery of additions to knowledge.

The major difficulties in the road to the formulation of useful hypothesis are three. First among these is the absence of (or the absence of knowledge of) a clear theoretical framework. Second is the lack of ability to utilize that theoretical framework logically.

Third is the failure to be acquainted with available research techniques so as to be able to phrase the hypothesis properly. These obstacles will be treated later, but at the moment it is possible to stop and consider the question, "Just how difficult is it to ask an important, testable question?"

Let the student answer this question himself. By the time he completes the course, he will have had several sociology courses. If he happens to be a superior student, he has also read several monographs in sociology. With this knowledge of sociological theory at hand, let him formulate one *good, definite, testable hypothesis*.

Many students will completely fail such a test. If so, they should not be discouraged for, this is not a simple task. In any case, one of the functions of this course is to improve the ability to formulate good hypothesis. If the student is able to formulate propositions at all, closer investigation will show many of them not to be hypothesis. Some students will have merely selected an *area* of study: the socialization of the child, juvenile delinquency, white-collar crime, or courtship behaviour. Such formulations, at course, are not hypothesis; they do not formulate precise and testable questions.

Somewhat closer to the mark will be some who might suggest the *replication* of previous studies. That is, some may think it is useful to repeat a previous piece of scientific work, duplicating the conditions exactly. This is useful work and does in one sense state a hypothesis i.e., that the results will be the same. But the utility of this procedure does not go beyond checking findings and it is likely to make no contribution to new knowledge.

Still closer to the formulation of a hypothesis would be those few who might suggest the study of *empirical regularities*. This type of research would be represented by the study of such things as the ecological distribution of mental disorders, the acceptance of contraceptive practices in Latin America, or the marital adjustment of rural Southerners. Such questions do suggest the *type of data* to be gathered, but they are hypothesis of a low level of abstraction; they merely predict that *some* type of patterning will appear without predicting *what* that pattern will be.

On the other hand, if we actually begin with a broad theory, and by deduction predict a social regularity as a relationship between two or more factors, we may develop a hypothesis. We might then obtain such formulations as the following, although space does not allow a statement of the entire chain of theoretical reasoning upon which they are based or the detailed definitions necessary:

### **1. Principle**

A socially recognized relationship in which there are strains built into the situation will also be surrounded by institutionalized controls, to ensure conformity of the participants with implicit or explicit norms.

#### ***Deduction***

We therefore predict that in those professions (such as psychiatry and psychotherapy generally, medicine, and law to a lesser degree) which deal with the more intimate aspects of clients' lives there are (a) more emotional strains in the client-practitioner relationship, and (b) more internalized and external controls upon both participants than is the case in other professions (such as engineering, architecture, dentistry). Of course, such a hypothesis can and must be broken down into subhypotheses. These would take these forms: (a) specification of the *degree* of difference; (b) specification of profession and problem, to separate criminal law from corporation law, types of contacts between profession and client, and types of strain-producing problems; and (c) specification of kinds of controls.

### **2. Principle**

Rather extensive, but relatively unsystematized, data show that members of the upper occupational-class strata experience less unhappiness and worry and are subject to more formal controls than members of lower strata.

#### ***Deduction***

Our hypothesis would then predict that this comparison also applies to the marital relationships of members of these strata and

would predict that such differential pressures could be observed through divorce rates. There should be an inverse correlation between class position and divorce rates. Again, we would have to define our terms carefully and show the systematic connection between our original premises and our deduction, but the result can be tested by the degree of our correlation.

The above examples indicate not only the difficulty of formulating a hypothesis, but also the need to do so. Early in any investigation a definite hypothesis should be formed. At first this may not be very specific. In such an instance it is referred to as a "working hypothesis," which will be subjected to modification as the investigation proceeds. The utilization of a hypothesis, however, is necessary for any useful research results.

### TYPES OF HYPOTHESIS

What are the *kinds* of hypothesis with which the sociologist deals? There are many ways of classifying hypothesis. For the purpose of this book, however, it seems adequate to separate them on the basis of the level of abstraction. Three broad levels may be distinguished. These will be discussed in the order of increasing abstractness.

**Some hypothesis state the existence of empirical uniformities.** These hypothesis frequently, though not always, represent the scientific examination of common-sense propositions. Thus, we might make a survey of some area that seems to represent a "problem" in common-sense terms. It usually represents, also, a problem about which some "common-sense" observations already exist. There are many types of such empirical uniformities which are common in sociological research. These studies may show regularities in the distribution of business establishments in a city, the ethnic backgrounds of workers in an industry, the size of families on relief, or the distribution of Negroes in a nation.

Or, they may describe the behaviour patterns of specific groups—for example, the students at a particular college in their freshman year. Here we might tabulate conformity and nonconformity to

customary usage; the wearing of the "dink," the submission to "initiation" rites, or the pledging to a fraternity. From research of this type the tabulations will yield expressions of the *degree of uniformity* in social behaviour. They may be symbolized by graphs, figures, or maps. In any event, their end product is a simple description of group activities.

It may be protested, of course, that these investigations do not involve the *testing* of a hypothesis at all, but are merely adding up the facts. Such a charge may have merit in a particular case, but the line is difficult to draw. Certainly, many such studies have actually sought to test common-sense statements about these phenomena, using such statements as hypothesis. It may be further objected that these are not useful hypothesis, on the grounds that they merely represent what everyone already knows. There are two answers to this objection. First, "what everyone knows" is not put in precise terms nor is it integrated into the framework of science. Second, "what everyone knows" may well be incorrect. To put commonsense ideas into precisely defined concepts and subject the proposition to test is an important task of science.

These statements are particularly true for sociology at its present, early stage of development. Folk knowledge of social relations is abundant, but it is often a confused mixture of cliches and moral judgments. Sociology thus has a large-scale job in transforming and testing these so that they can become useful knowledge. This requires that three tasks be performed: first, the removal of value judgments; second, the clarification of terms; and third, the application of validity tests.

For example, such statements as "Bad children are born that way" or its reverse, "Bad parents produce bad children," or "Wealthy people have a high divorce rate because they lead such self-indulgent lives" are the kinds of generalizations which, though commonplace, cannot be tested. As they stand, they merely express sentiment rather than describe fact, and the concepts are unclear. They could be made into adequate hypothesis, however, if cleared of moral overtones and put into carefully defined terms.

Not only sociology, as noted previously, but all science has found such common-sense knowledge fruitful for study even when it has been wrong. "Everybody knew" that the sun revolved around the earth, that horsehairs in a watering trough would turn to worms, that a bag of asafoetida hung around the neck would prevent colds (this last, at least, may have been true, since the smell kept others at a distance!). All those beliefs have been exploded by patient, plodding empirical checking of the facts, so that we now know that horsehairs do not turn into worms and asafoetida has no effect on colds.

In social relations, too, there are many cliches which are not correct. The objection that it "elaborates the obvious" has been made by a good number of critics against the monumental work *The American Soldier* (Princeton, N.J.: Princeton University Press, 1949-1950; 4 vols.). It would seem, for example, that there was no need for the social researchers to prove the following hypothesis, since they were known already:

1. Soldiers from white-collar jobs were somewhat less adjusted in the Army, since they had sacrificed more than lower class men by going into the service.
2. Negro soldiers, knowing that the barriers against promotion were rigid, did not work for promotion so hard as did white soldiers.
3. Soldiers in units with high promotion rates had a more optimistic view of promotion chances and were more satisfied about promotion policies than were soldiers in units with low promotion rates.

Nevertheless, these were among the hypothesis tested with the result that all three were proved to be *incorrect*. Often, we believe that "everybody knows that," but we make the statement *after* the investigation. We could not have predicted the result. We believe that the result is only common sense, since *some* of our experience fits the result. However, if the result had turned out differently, we would have found still other experiences, of a contrary

order, to fit the different results. As a consequence, many supposedly obvious facts must actually be tested. It hardly needs to be added, moreover, that even when we know in general that a given relationship exists, we do not know to what degree or in what proportions exists. Science demands a higher precision than "in general".

In any case it is certain that "what everybody knows" is not known until it has been tested. The simple level of hypothesis that seeks empirical generalization plays an important role in the growth of science.

**Some hypothesis are concerned with complex ideal types.** These hypothesis aim at testing the existence of logically derived relationships between empirical uniformities. If this test sounds difficult to understand, an example may help to make it clearer. Human ecology early described a large number of empirical uniformities. Land values, industrial concentrations, types of businesses, ethnic groups, mental disorders, and many other phenomena appeared to show unquestionable uniformities in distribution. Further study and logical analysis of these and other related findings led to the formulation of various hypothesis concerning the way in which these were related. One such hypothesis was Ernest W. Burgess's statement of the concentric growth circles that characterize the city.

This hypothesis was then tested against a variety of variables in a number of cities. That this ideal type does represent the actual pattern of city growth is not accepted by all ecologists, however, and so this formulation remains a hypothesis until a more crucial test of it is made.

Another hypothesis concerning an ideal type resulted from these same ecological empirical uniformities. This was the notion that areas tend to represent certain characteristics in a series of predictable patterns. This was called the hypothesis of "the natural area." Much research has been done on this hypothesis, and the results, although they have modified the original statement somewhat, have generally supported it. With the growth of supporting evidence, notions about natural areas have become a part of sociological theory

rather than remaining hypotheses. A similar type of hypothesis in another area resulted from the analysis of minority groups. Many studies revealed empirical uniformities in the behavior of members of a wide variety of minorities. Logical analysis then led to the hypothesis that these uniformities produced an ideal type. This was at first called by H. A. Miller the "oppression psychosis," but it was subsequently modified to the "marginal man" by E. W. Stonequist and others. Empirical evidence supported the hypothesis, and thus the "marginal man" is today also a part of sociological theory. It is important to see here that this level of hypothesizing moves beyond the expectations of simple empirical uniformity, by *creating* a complex referent in society. Not *all* areas must be natural areas, not *all* members of minority groups must be marginal men, not *all* cities must show perfect concentric circles, for these hypotheses to be useful. They must, of course, be verified in that under *certain conditions* of maximum opportunity such instances will occur, but in reality such hypotheses are purposeful distortions of empiric exactness. Because of their removal from empirical reality these constructs are termed "ideal types." The function of such hypotheses is to create tools and problems for further research in otherwise very complex areas of investigation.

**Some hypothesis are concerned with the relation of analytic variables.** These hypothesis occur at a level of abstraction beyond that of ideal types. Whereas the hypothesis of empirical uniformities lead to the observation of simple differences, and those dealing with ideal types lead to specific coincidences of observations. The study of *analytic variables* requires the formulation of a relationship between changes in one property and changes in another.

To take an example from sociology, the study of human fertility might show empirical regularities by wealth, region, size of community, and religion. If this were then raised to the level of ideal type formulation, one result might be the hypothesis that there are two high-fertility population segments in the United States. One would be the low-income, Southern, rural Protestant, and the other the low-income, Northern, urban Catholic. At a still higher level of

abstraction the *qualities* of region, size of community, and religion might be abstracted and controlled; that is, their effects on fertility held constant. This would allow a better measurement of the relation between the *variables* of wealth and fertility. Similarly, the problem could be stated in such a way that any three could be controlled so as to allow the fourth to vary and hence to measure its relation to fertility. It is clear that this is a very abstract way to handle the problem because *there are no people* whose fertility is not affected by all the variables. Of course, not all the characteristics mentioned are as yet expressed as variables.

This level of hypothesizing is not only more abstract than the others; it is also the most sophisticated and the most flexible mode of formulation. At this level, the number of variables which can be abstracted and studied is limited only by theory; and since theory grows by the process itself, opportunities for new research are constantly being created.

In the event that it should appear that any of these types of hypotheses is "better" than another, a word of explanation may be needed. The function of the ideal-type method, it will be recalled, is to provide constructs for use in further hypothesizing. This is also one function of studies of empirical uniformities. Without the painstaking, grubbing labor which characterizes this type of investigation, none of the "brilliant" theories of a more abstract nature could have ever appeared. Particularly in sociology is it necessary for the student to learn that at whatever the level of abstraction the hypothesis lies, the need for careful work does not vary, nor is the significance of the findings automatically apparent.

Thus far in the chapter, three major points have been made: (1) that a hypothesis is a necessary condition for successful research; (2) that formulation of the hypothesis must be given considerable attention, to clarify its relation to theory, remove vague or value-judgmental terms, and specify the test to be applied; and (3) that hypotheses may be formulated on different levels of abstraction. At various points in the discussion, more or less casual references have been made to the question of the *origins* of hypotheses. At

this point it seems useful, then, to look at this matter in greater detail and somewhat more systematically. It seems possible to distinguish four such sources more or less clearly.

**General culture in which a science develops furnishes many of its basic hypothesis.** This point has been mentioned several times before in the discussions of science and values, pure and applied science, and the simplest hypotheses which state empirical regularities. It has been pointed out that science has developed in Western society and that this is no mere accident but is a function of the culture itself.

The fact that sociology is so new and that its growth has taken place very largely in the United States, England, Germany, and France means that the hypotheses which have been put forth and tested have been related to a particular cultural complex. To oversimplify the situation, let us assume that the American variant of Western European culture emphasizes individual happiness, mobility, and competition. This is in contrast, let us say, to the Zuni type of Pueblo culture in which there is more emphasis upon the group, an avoidance of personal competition and achievement, and less concern about individual happiness. Flowing from this, certain hypotheses could be expected to occupy the attention of American sociologists. To say that these hypotheses are the product of the cultural values does not make them scientifically less important than others, but it does at least indicate that attention has been called to them by the culture itself.

For example, the American emphasis upon personal happiness has had considerable effect upon social science in this country. Not only is there an excellent market for books explaining "how to be happy," but the phenomenon itself has been studied in great detail. Much of textbook economics is based upon a theory of human action which is predicated upon personal happiness as the central motivation. There have been many studies of the factors which make for marital happiness. Even the term "adjustment" used by sociologists and psychologists customarily means happiness and is the focus of innumerable studies. Happiness in one way or another

has been correlated with income, education, occupation, ethnic origin, social class, and parental happiness. The factors contributing to adjustment in sexual relations, marital relations, on the job, and in other social groups have been analyzed in detail. From all this it is at least clear that the cultural emphasis upon happiness has been productive of an almost limitless range of hypotheses for American social science.

Not only do the major cultural values serve to direct research interests, folk wisdom serves as another source of hypotheses. In Western society, in varying degrees, race is thought to be an important determinant of human behavior. This is perhaps most widely and extremely held in the United States and in South Africa. The sociologist in this cultural setting cannot accept such a folk belief as fact but must test it scientifically. It would be very simple to think of an almost limitless number of similar common-sense propositions which have served or could serve as a source of hypotheses.

This raises still another point. It is not merely that the existence of such propositions is productive of hypotheses but also that social change increases the value of the culture as a source of scientific questions. Common-sense propositions are usually unquestioned. Ideas and behavior often seem so obvious as to call for no serious study. It is, in these cases, a real test of the investigator's ingenuity to see a question in such truisms. Social change, however, may call these into question, thus providing a hypothesis for study. Thus, doctrines of both "liberalism" and "progressivism" have played important roles in social science. The latter, by embracing change, challenges the old assumptions; and the former, by emphasizing the importance of the individual, insists that he not be pre-judged. In either case there is present the kind of skepticism which is productive of hypotheses.

For example, the folk notions about race were called into question on *moral* grounds. The progressive and liberal ideology held the old notion of the racial determination of behavior to be false. Careful analysis of the Army Alpha tests of World War I, studies of the IQs of Negroes and Whites, anthropological evidence

about the learning ability of "primitives," and many other studies piled up scientific evidence opposed to the older folk beliefs. Similar results occurred in other areas, such as the inherent lack of capabilities in the lower classes. Alcoholism is no longer considered to be the result of weak moral fiber but is regarded as a disease. These examples could be multiplied almost indefinitely, but enough have already been given to indicate the role of "equalitarian" thought patterns and of social change, in the generation of hypotheses.

**Hypothesis originate in the science itself.** Mention has already been made of the fact that this operates in two ways. First, in the discussion on theory and fact it was pointed out that theory gives direction to research by stating what is known. Logical deduction from this leads to new problems. Second, in the treatment of values and science it was pointed out that science is a social relation and that the scientist must acquire the folkways of his discipline.

As an example of the way the first effect comes about, a development in communications research may be cited. It was first established in theory ???? that there existed people who could be considered as "opinion leaders. It was further seen that these were prestigious people, that is, that they possessed high status in the community. Since it was also known that high status is a function of a number of variables, it was logical to hypothesize the existence of an ideal type, "the influential person." However, in an actual study, the influential persons did not seem to have many characteristics in common. This led to still further questioning and the development of two major categories, which Merton termed "cosmopolitan" and "local" influentials, each with its own set of characteristics. Thus, what was known led to the asking of still another question or, in other words, to the formulation of a new hypothesis. Indeed, the student will find a number of suggested readings at the end of this chapter, which deal with hypotheses whose origins lay in the *deviant* cases.

The "socialization" process in learning a science also affects the hypotheses which will be developed by the scientist. First of all the student learns from his teachers which are the promising areas,

which methods are adequate, which scholars are superior, and, of course, which are "inferior." Thus, the range of hypotheses open to him is limited by the direction of his learning experience. Later in his life, the scientist is affected by a similar process—the approval of his colleagues. Formally and informally, scientists continually discuss current research, both orally and in print. In this way consensus is reached as to which areas and problems are thought to be important. Through this constant interaction in the area, "fashionable" modes of thought, terminology, concepts, and problems develop. These, in turn, of course, operate to suggest further hypotheses.

**Analogy is often a source of useful hypothesis.** Julian Huxley has pointed out that casual observations in nature or in the framework of another science may be a fertile source of hypotheses. The hypotheses that resulted in the development of human ecology, for example, were an application of established theory in the fields of plant and animal ecology. Thus, the hypothesis that similar human types or activities may be found occupying the same territory came from plant ecology, where the phenomenon is known as segregation. When the hypothesis was borne out by social observation, the same term was taken into sociology where it has become an important idea in sociological theory.

Similarly, the observation that the behavior of human groups seems to exhibit some of the same patterns as found in gravitational and electric fields led to the basic hypothesis of what is called social physics. This hypothesis is that if people are related in some way similar to the structure of such a physical field then human behavior should show reliable correlation with the values secured by such a field analysis. John Q. Stewart has now published much evidence to indicate that the application of this analogy might be interesting.

The use of analogy as a source of hypotheses is not without its dangers, of course. There is reason to suspect any analogy from another science, since the models to be applied are clearly understood in their own theoretical framework but are not related to the new frame of reference. Thus, it is dangerous to assume that natural areas in human society are a product of symbiosis as is true in

biology. We have no empirical method of applying the concept to human beings. Similarly, it is dangerous to assume, as in social physics, that "demographic potential" is the same phenomenon as "gravitational potential" in physics. In short, analogy may be very suggestive, but care must be taken not to accept models for sociology from other disciplines without careful examination of the concepts which make up the models.

Hypothesis are also the consequence of personal, idiosyncratic experience. Not only do culture, science, and analogy affect the formulation of hypotheses. The way in which an *individual* reacts to each of these is also a factor in the statement of hypotheses. Therefore, the individual experience of the scientist contributes to the type and the form of the questions he asks.

In just the same way that perception has been shown to be structured by experience, producing odd and interesting illusions, some persons will perceive an interesting pattern from what may merely seem a jumble of facts to another. The history of science is full of instances of discoveries made because the "right" individual happened to make the "right" observation because of his particular life history.

Thomas Henry Huxley is reported to have exclaimed, on reading Darwin's *Origin of Species*, "Oh, what an ass I was not to have thought of that!" Even if the story is apocryphal, it is pertinent, for Darwin had assembled many facts which had been known for at least two generations. In addition he had added many observations on his famous 4-year voyage on H.M.S. *Beagle*. Nevertheless, this enormous body of data did not take on any systematic order until a fortuitous event occurred. Darwin was pondering the problem of understanding what caused species to change, when he happened to read Thomas Malthus's notion that a population tended, in the absence of certain other checks, to overwhelm the resources for the sustenance of that population. In other words, the physical environment itself was always snapping at the heels of any species. If individuals change in certain directions they will be at an advantage; if they change in other directions, at a disadvantage. This, then,

combined with Darwin's other information, resulted in the notion of the struggle for survival of the species. After its public expression and in spite of the fierce theological controversy it aroused, this explanation was quickly accepted by scientists. Huxley was simply exclaiming, because "anyone could have seen it." This was indeed the "right" man at the "right" time.

This should not be construed that certain ideas will be observable only by one particular man. In fact, Wallace independently worked out the same idea as Darwin but decided the latter's greater body of data justified publication by him. All discoveries are made not once but many times. It is merely that personal life histories are a factor in determining the kinds of perception and conception. These factors may, in turn, direct one person to certain hypotheses more quickly. Often, of course, these persons were not seeking the particular observation or hypothesis. They were simply trained to understand and use the strategic fact when it appeared. The story of Newton and the falling apple, however untrue, illustrates this individual, accidental process. Similar occurrences are by no means unknown in the scientific laboratory.

An illustration of individual perspective in the social sciences may be seen in the work of Thorstein Veblen. The product of an isolated Norwegian community in Minnesota, Veblen lived at a time when the capitalistic system was not usually called into question except by "radicals." His own community background, however, was replete with negative experiences concerning the working of the economic system, and he was himself a kind of outsider or "marginal man," able to look at the capitalist system objectively. He was thus in an excellent position to attack the fundamental concepts and postulates of classical economics. In a very real sense he was an alien who could bring a different experience to bear upon the economic world. As a result he made penetrating analyses of our society which have profoundly influenced social science since his time.

All these sources of hypotheses value orientations of the culture, folk wisdom and cliche, rebellion against common-sense ideas,

observation of deviant cases (the cases which "don't fit the rule"), social experience within the science, the application of analogies, and personal experience provide a wealth of hypotheses. In fact it is an almost embarrassing profusion. The problem which this raises is how to select those ideas which may actually prove useful.

## CHARACTERISTICS OF USABLE HYPOTHESIS

In the privacy of the scientist's mind, alone or in social gatherings, in odd moments or in the press of business, many hypotheses are entertained. Most of them, having appeared, are fortunately left to die alone. A few survive, however, to be exhibited in "bull sessions" or to be tried out on sleepy undergraduates at eight o'clock on a wintry morning. Most are not destined to play any significant role in the growth of science. Some of these seem to be the product of the fact, as the philosopher Suzanne *Langer* has argued, that man's mind, like his body, is often active without any immediate goal. It is only by the imposition of firm standards that it is possible to winnow out the good ideas from the bad.

Let us now look at some criteria for judging hypothesis.

### Hypothesis must be conceptually clear

Enough emphasis upon this requirement was made in the preceding chapter to require little further elaboration. It should be repeated, however, that this involves two things. The concepts should be clearly defined, operationally if possible. Moreover, they should be definitions which are commonly accepted and communicable rather than the products of a "private world."

### What to do

One simple device for clarifying concepts is to write out a list of the concepts used in the research outline. Then try to define them (a) in words, (b) in terms of particular operations (index calculations, types of observations, etc.), and (c) with reference to other concepts to be found in previous research. Talk over each concept with fellow students and other researchers in the field. It will often be found that supposedly simple concepts contain many

meanings. Then it is possible to decide which is the desired referent. For systematic conceptual clarification, perform all the operations suggested in Chapter 5.

### Hypothesis should have empirical referents

It has also been previously pointed out that scientific concepts must have an ultimate empirical referent. No usable hypotheses can embody moral judgments. Such statements as "criminals are no worse than businessmen," "women should pursue a career," or "capitalists exploit their workers," are no more usable hypotheses than the familiar proposition that "pigs are well named because they are so dirty" or the classical question, "How many yards of buttermilk are required to make a pair of breeches for a black bull?" In other words, while a hypothesis may study value judgments, such a goal must be separated from a moral preaching or a plea for acceptance of one's values.

### What to do

First, analyze the concepts which express attitudes rather than describing or referring to empirical phenomena. Watch for key words such as "ought," "should," "bad," etc. Then transform the notions into more useful concepts. "Bad parents" is a value term, but the researcher may have a definite description in mind: parents who follow such practices as whimsical and arbitrary authoritarianism, inducing psychic insecurity in the child, failure to give love, etc. "Should" is also a vague term, but the student may simply mean, "If women do not pursue a career, we can predict emotional difficulties when the children leave home, or we can predict that the society will not be able to produce as much goods," etc. When, instead, we find that our referent is simply a vague feeling, and we cannot define the operations needed to observe it, we should study the problem further and discover what it is that we really wish to investigate.

### Hypothesis must be specific

That is, all the operations and predictions indicated by it should

be spelled out. The possibility of actually testing the hypothesis can thus be appraised. Often hypotheses are expressed in such general terms, and with so grandiose a scope, that they are simply not testable. Because of their magnitude, such grand ideas are tempting because they seem impressive and important. It is better for the student to avoid such problems and instead develop his skills upon more tangible notions.

By making all the concepts and operations explicit is meant not only conceptual clarity but a description of any *indexes* to be used. Thus, to hypothesize that the degree of vertical social mobility is decreasing in the United States requires the use of indexes. At present there is no satisfactory operational definition of the status levels which define mobility. Therefore, the hypothesis must include a statement of the indexes which are to be used; that is, political office, occupation, effective income, education, etc.

Such specific formulations have the advantage of assuring that research is practicable and significant, in advance of the expenditure of effort. It furthermore increases the validity of the results, since the broader the terms the easier it is to fall into the trap of using *selective evidence*. The fame of most prophets and fortunetellers lies in their ability to state predictions so that almost any occurrence can be interpreted as a fulfillment. We can express this in almost statistical terms: the more specific the prediction, the smaller the chance that the prediction will actually be borne out as a result of mere accident. Scientific predictions or hypotheses must, then, avoid the trap of selective evidence by being as definite and specific as possible.

### **What to do**

Never be satisfied with a general prediction, if it can be broken into more precise subhypotheses. The general prediction of war is not enough, for example: we must specify time, place, and participants. Predicting the general decline of a civilization is not a hypothesis for testing a theory. Again, we must be able to specify and measure the forces, specify the meaning and time of decline, the population

segments involved, etc. Often this can be done by conceptual analysis and the formation of related hypotheses; e.g., we may predict that urbanization is accompanied by a decline in fertility. However, we gain in precision if we attempt to define our indexes of urbanization; specify which segments will be affected, and how much (since in the United States the various ethnic and religious segments are affected differently); specify the amount of fertility decline, and the type (percentage childless, net reproduction rate, etc.). Forming sub-hypotheses (1) clarifies the relationship between the data sought and the conclusions; and (2) makes the specific research task more manageable.

### **Hypothesis should be related to available techniques**

It has already been pointed out that theory and method are not opposites. The theorist who does not know what techniques are available to test his hypotheses is in a poor way to formulate usable questions.

This is not to be taken as an absolute injunction against the formulation hypotheses which at present are too complex to be handled by contemporary technique. It is merely a sensible requirement to apply to any problem in its early stages in order to judge its researchability.

There are some aspects of the impossible hypothesis which may make its formulation worth while. If the problem is significant enough as a possible frame of reference it may be useful whether or not it can be tested at the time. The socioeconomic hypothesis of Marx, for example, were not proved by his data. The necessary techniques were not available either then or now. Nevertheless, Marxian frameworks are an important source of more precise, smaller, verifiable propositions. This is true for much of mile Durkheim's work on suicide. His related formulations concerning social cohesion have also been useful. The work of both men has been of paramount importance to sociology, even though at the time their larger ideas were not capable of being handled by available techniques.

Further more, posing the impossible question may stimulate the growth of technique. Certainly some of the impetus toward modern developments in technique has come from criticisms against significant studies which were considered inadequate because of technical limitations. In any serious sociological discussion, research frontiers are continuously challenged by the assertion that various problems "ought" to be investigated even though the investigations are presently impossible.

### **What to do**

Look for research articles on the subject being investigated. Make a list of the various techniques which have been used to measure the factors of importance in the study. If you are unable to locate any discussions of technique, you may find it wiser to do a research 071 *the necessary research techniques*. You may, instead, decide that this lack of techniques means your problem is too large and general for your present resources.

Some items, such as stratification or race attitudes, have been studied by many techniques. Try to discover why one technique is used in one case and not in another. Note how refinements in technique have been made, and see whether one of these may be more useful for your purposes. Look for criticisms of previous research, so as to understand the weaknesses in the procedures followed.

Again, other problems may have been studied with few attempts at precise measurement. Study the literature to see why this is the case. Ascertain whether some subareas (for example, of religious behavior) may be attacked with techniques used in other areas (for example, attitude measurement, stratification measures, research on choice making, etc.).

### **Hypothesis should be related to a body of theory**

This criterion is one which is often overlooked by the beginning student. He is more likely to select subject matter which is "interesting", without finding out whether the research will really help to refute, qualify, or support any existing theories of social relations. A science,

however, can be cumulative only by building on an existing body of fact and theory. It cannot develop if each study is an isolated survey.

Although it is true that the clearest examples of crescive theoretical development are to be found in the physical and biological sciences, the process can also be seen in the social sciences. One such case is the development of a set of generalizations concerning the social character of intelligence. The anthropological investigations at the end of the nineteenth century uncovered the amazing variety of social customs in various societies, while demonstrating conclusively that there were a number of common elements in social life: family systems, religious patterns, an organization of the socialization process, etc.

The French school of sociology, including Lucien Levy-Bruhl, Smile Durkheim, Marcel Mauss, Henri Hubert, and others, formulated a series of propositions, at the turn of the century, which suggested that the intellectual structure of the human mind is determined by the structure of the society. That is, perception and thought are determined by society, not alone by the anatomical structure of our eyes, ears, and other senses. Modes of thought vary from society to society. Some of these formulations were phrased in an extreme form which need not concern us now, and they were often vague. Nevertheless, the idea was growing that the intelligence of a Polynesian native could not be judged by European standards; his thinking was qualitatively, not merely quantitatively, different.

At the same time, however, better techniques were being evolved for measuring "intelligence," which came to be standardized in the form of scores on various IQ tests. When these were applied to different groups it became clear that the variation in IQ was great; children of Italian immigrants made lower grades on such tests, as did Negroes. Northern Negroes made higher grades than whites from many Southern states. American children of Chinese and Japanese parents made rather high scores. Since it was generally assumed that these tests measured "innate intelligence," these data were sometimes generalized to suggest that certain "racial" groups

were by nature inferior and others superior. However, such conclusions were opposed on rational grounds, and liberal sentiments suggested that they be put to the test. There were, then, two major sets of conclusions, one suggesting that intelligence is in the main determined by social experience, the other suggesting that the *IQ* is innately determined. To test such opposing generalizations, a research design was needed for testing logical expectations in more specific situations. If, for example, it is true that the intelligence of individuals who are members of "inferior" groups is really determined biologically, then changes in their environments should not change their IQ. If, on the other hand, the social experience is crucial, we should expect that such changes in social experience would result in definite patterns of IQ change. Further deductions are possible. If identical twins are separated and placed in radically different social experiences at an early age, we might expect significant differences in IQ. Or, if a group of rural Negro children moves from the poor school and social experience of the South, to the somewhat more stimulating environment of the North, the group averages would be expected to change somewhat. Otto Klineberg, in a classic study, carried out the latter research. He traced Negro children of various ages after they had moved to the North and found that, in general, the earlier the move to the North occurred, the greater the average rise in the IQ. The later the move, the smaller the increase. Even if one assumes that the "better," more able, and more daring adult Negroes made this move, this does not explain the differences by time of movement. Besides, of course the subjects were children at the time of the migration.

In this research design a particular result was predicted by a series of deductions from a larger set of generalizations. Further, the prediction was actually validated. In justice to the great number of scholars who have been engaged in refining and developing IQ tests, it should be mentioned that other tests and investigations of a similar order have been carried out by many anthropologists, sociologists, and social psychologists. They do not invalidate the notion that IQ is based in part on "innate" abilities, but they do

indicate that to a great extent these abilities must be stimulated by certain types of experience in order to achieve high scores on such tests.

From even so sketchy an outline of a theoretical development as the foregoing is, it can be seen that when research is systematically based upon a body of existing theory, a genuine contribution in knowledge is more likely to result. In other words, to be worth doing, a hypothesis must not only be carefully stated, but it should possess theoretical relevance.

### ***What to do***

First, of course, cover the literature relating to your subject. If it is impossible to do so, then your hypothesis probably covers too much ground. Second, try to abstract from the literature the way in which various propositions and sets of propositions relate to one another (for example, the literature relating to Sutherland's theory of differential association in criminology, the conditions for maximum morale in factories, or the studies of prediction of marital adjustment). Third, ascertain whether you can deduce any of the propositions, including your own hypothesis, from one another or from a small set of major statements.

Fourth, test it by some theoretical model, such as Merton's "Paradigm for Functional Analysis in Sociology" (*Social Theory and Social Structure*, pp. 50-54), to see whether you have left out major propositions and determinants. Fifth, especially compare your own set of related propositions with those of some classic author, such as Weber on bureaucracy or Durkheim on suicide. If you find this task of abstraction difficult, compare instead with the propositions of these menas explained by a systematic interpreter such as Talcott Parsons in his *Structure of Social Action*. What is important is that, whatever the source of your hypothesis it must be logically derivable from and based upon a set of related sociological propositions.

## HYPOTHESIS : ITS NATURE AND ROLE IN SOCIAL RESEARCH

Once the problem to be answered in the course of research is finally instituted, the researcher may proceed to formulate tentative, solutions or answers to it. These proposed solutions or explanations are called hypothesis, which the researcher is obliged to rest on the basis of facts already known or which can be made known. When such hypothesis are not formulated, even implicitly, the researcher cannot effectively go ahead with the investigation of his problem because, in the absence of direction which hypothesis typically provide, the researcher would not know what facts to look for and what relation or order to search for among them. The hypothesis guide the researcher through a bewildering jungle of facts to see and select only those that are relevant to the problem or difficulty he proposes to solve. Collection of facts merely for the sake of collecting them will yield no fruits. To be fruitful, one should collect such facts as are for or against some point of view or proposition. Such a point of view or proposition is the hypothesis. The task of the inquiry or research is to test its accord with facts.

Lundberg aptly remarks: "The only difference between gathering data without a hypothesis and gathering them with one is that in the latter case, we deliberately recognize the limitations of our senses and attempt to reduce their fallibility by limiting our field of investigation so as to prevent greater concentration of attention on particular aspects which past experience leads us to believe are insignificant for our purpose.

Simply stated, a hypothesis helps us and appreciate (1) the kind of data that need be collected in order to answer the research question and (2) the way in which they should be organized most efficiently and meaningfully.

Webster's New International Dictionary of English Language, 1956, defines the term "hypothesis" as a "proposition, condition or principle which is assumed, perhaps without belief, in order to draw out its logical consequences and by this method to test its accord with facts which are known or may be determined."

Cohen and Nagel bring out the value of hypothesis thus: "We cannot take a single step forward in any inquiry unless we begin with a suggested combination or solution of the difficulty which originated it. Such tentative elpinations are suggested to us by something in the subject-matter and by our previous knowledge. When they are formulated as propositions, they are called hypothesis."

Once the scientist knows what his question (problem) is, he can make a guess, or a number of cuesses as to its possible answers. According to Werkmeister, "The guesses he makes are the hypothesis which either solve the problems or guide him in further investigation."

A hypothesis is a provisional formulation; a tentative solution of the problem posed by the scientist. The scientist starts by assuming that the solution is true without, of course, personally believing in its truthfulness. Based on this assumption, the scientist anticipates that certain logical consequences will be observed on the plane of observable events or objects. Whether these anticipalions or expectations really materialize is the test of the hypothesise its proof or disproof. When the hypothesis is proved, the problem of which it was tentative solution is answered. If it is not proved, i.e., falsified owing to non-support of proof, alternative hypothesis or solutions would be required to be formulated and tested. A hypothesis thus stands somewhere at the midpoint of research; from here one can look back to the problem and also look forward to data. The hypothesis may be slated in the form of a principle, that is, the tentative explanation or solution to the question how? or why? may be presented in the form of a principle that X varies with Y. If the inquiry establishes that an empirical referent of X varies with the empirical referent of Y in a concrete observably situation, (i.e., the hypothesis is proved) then the question is answered. Hypothesis, however, may take other forms, such as intelligent guesses, conditions, propositions deduced from theories, etc.

Finding on the basis of hypothesis are slow and hard way of science. While some scientific conclusions and premises seem to have arisen in the mind of the investigator as if by flashes of insight, in a majority of cases the process of discovery has been a slower

one. "The scientific imagination devises a possible solution a hypothesis and the investigator proceeds to test it. He makes intellectual keys and then tries whether they fit the lock. If the hypothesis does not fit, it is rejected and another is made. The scientific workshop is full of discarded keys.

Cohen and Nagel's statement that one cannot take a single step forward in any inquiry, without a hypothesis may well be an accurate statement of the value of hypothesis in scientific investigation generally, but it hardly does justice to an important function of scientific research, i.e., the "formulation of hypothesis." Hypothesis are not given to us ready made. Ofcourse in fields with a highly developed theoretic structure it is reasonable to expect that most empirical studies will have at least some sharp hypothesis to be tested. This is so not especially in social sciences where there not yet evolved a highly developed theoretical system in many areas of its subject-matter which can afford fruitful bases for hypothesis formulation. Attempts to force research into this mould are either deceitful or stultifying and hypothesis are likely to be no more than hunches as to where to look for sharper hypothesis in which case the study may be described as an intelligent fishing trip. As a result, in the social sciences at least, a considerable quantum of research endeavour is directed understandably toward 'making' hypotheses rather than at testing them. As we shall see in alater chapter, a very important type of research has as its goal, the formulation of significant hypothesis relating to a particular problem. Hence, we will do well to bear in mind that research can begin with well formulated hypothesis or it may come out with hypothesis as its end product.

The role of hypothesis for research in the words of Chaddock who summarizes it thus: "(A hypothesis) in the scientific sense is an explanation held after careful canvasss of known facts, in full knowledge of other explanations that have been offered and with a mind open to chaise of view, if the facts disclosed, by the inquiry warrant a different explanation. It is therefore held with the definite purpose of including in investigation all available and pertinent data either to prove or disprove the hypothesis.... (A hypothesis) gives

point to the inquiry and if found on sufficient previous knowledge, guides the lines of investigations. Without it much useless data may be collected in the hope that nothing essential will be omitted or important data may be omitted which could have been easily included if 'the purpose of inquiry had been more clearly defined'; and thus hypothesis are likely to be no more than hunches as to where to look for sharper hypothesis in which case the study may be described an intelligent fishing trip.

### **VARIOUS KINDS OF HYPOTHESIS**

There are many kinds of hypothesis the social researcher has to be working with. One type of hypothesis asserts that something is the case in a given instance that a particular object, person or situation has ticular characteristic. Another type of hypothesis deals with the frequency of occurrences or of association among variables; this type of hypothesis may state that X and Y proportion of times or that urbanism tends to be accompanied by mental disease or that something is greater or lesser than some other thing. Yet another type of hypothesis asserts that particular characteristic is one of the factors which determine another characteristic, i.e X is the producer of Y (product). Hypothesis of this type are known as causal hypothesis.

Hypothesis can be classified in a variety of ways. But classification of hypothesis on the basis of their levels of abstraction is regarded as especially fruitful. Goode and Hatt have identified three differential levels of abstraction reached hypothesis, and hence typific. We shall then be starting from the lowest level of abstraction and go over to:

(a) At the lowest level of abstraction are the hypothesis which state existence of certain empirical uniformities. Many types of such empirical uniformities are common in social research, for instance it may be the sized with reference to India that in the cities men will get married between the age of 22 and 24 years. Or, the hypothesis of the type may state that certain behaviour patterns may be expected in a specified community. Thus, hypothesis of this type frequently

seem to invite scientific verification of what are called "common sense propositions," indeed without justification.

It has often been said by way of a criticism of such hypothesis that "these are not useful in as much as they merely state what everyone seems to know already. Such an objection may however be overruled by pointing out that what everyone knows is not often put in precise terms nor is it adequately integrated into the framework of science. Secondly, whatever everyone knows may well be mistaken. To put common sense ideas into precisely defined concepts and subject the proposition to test is an important tactic of science. This is particularly applicable to social sciences which are at present in their earlier stages of development." Not only social science but all sciences, have found such commonsense knowledge: a fruitful item of study. It was commonsense knowledge in the 'olden days' that 'the sun revolved round the earth. But this and many other beliefs based on commonsense have been exploded by patient, plodding empirical checking of facts. The monumental work, *The American Soldier* by Stouffer and associates was criticized in certain quarters, for it was according to them a mere elaboration of the obvious. But to this study goes the credit of exploding some of the commonsense propositions and shocking many people who had never thought that what was so obvious a commonsense could be totally wrong or unfounded in fact.

(b) At a relatively higher level of abstraction are hypothesis concerned with complex 'ideal types'. These hypothesis aim at testing whether logically derived relationships between empirical uniformities obtain. This level of hypothesizing moves beyond the level of anticipating a simple empirical uniformity by visualizing a complex referent in society. Such hypothesis are indeed purposeful distortions of empirical exactness and owing to their remoteness from empirical reality these constructs are termed "ideal types. The function of such "hypothesis is to create tools and formulate problems for further research in complex areas of investigation. Example of one such hypothesis may be cited. Analysis of minority groups brought to light empirical uniformities in the behaviour of

members of a wide variety of minorities.' It was subsequently hypothesized that these uniformities pointed to an 'ideal type'. First called by H.A. Miller the 'oppression psychosis', this ideal-typical construction was Subsequently modified as the 'Marginal man' by E. Stonequist and associates. Empirical evidence marshalled later substantiated the hypothesis, and so the concept of marginality has very much come to stay as a theoretic construct in social sciences, and as a part of sociological theory.

(c) One can now come to the class of hypothesis at the highest level of. This category of hypothesis is concerned with the relation obtaining amongst analytic variables. Such hypothesis are statements about how changes in one property (abstracted from a class of concrete objects) will effect another property, e.g., a statement: of relationship between education and social mobility or wealth and fertility. It is easy to see that this level of hypothesizing is not only more abstract compared to others; it is also the most sophisticated and vastly flexible mode of formulation

This does not mean, however, that this type of hypothesis is 'suilerlor' 'better' than the other types. Each type of hypothesis has its own importance depending in turn upon the nature of investigation and the level of development the subject has achieved. The sophisticated hypothesis of analytical variables owe much of their existence to the building blocks' contributed by the hypothesis at the lower orders of abstraction.

## SOURCES OF HYPOTHESIS

Hypothesis may be developed from a variety of sources. We examine here, some of the major ones.

(1) The history of science provides an eloquent testimony to the fact that personal and idiosyncratic experiences of the scientist contribute a great deal to the type and form of questions he may ask, as also to the kinds of tentative answers to these questions (hypothesis) that he might provide. Some scientists may perceive an interesting pattern in what may merely seem a jumble of facts to the common man. The 'history' of science is full of instances of

discoveries made just because the 'right' person happened to make the 'right' observation owing to his characteristic life-history and exposure to a unique mosaic of events. Personal life histories' are-a fact of in determining the kinds of a person's perception and conception and this factor may in turn direct him to certain hypothesis quite readily. An illustration of such individual perspectives, in social sciences may be seen in the work of Thorstein Veblen whom Merton describes as a sociologist with a keen eye for the unusual and paradoxical. A product of an isolated Norwegian community, Veblen lived at time when the capitalistic system was barely subjected to any criticism. His own community background was replete with deprivational experience attributable to the capitalist system. Veblen being an outsider, was able to look at the capitalist economist system more objectively and with impassionate detachment. Veblen was thus strategically positioned to attack the fundamental concepts and postulates of classical economics. He was an alien who could bring a different experience to bear upon the economic, world. Consequently, he made penetrating analyses of society and economy which have ever since profoundly influenced social science.

(2) Analogies are often a fountain head of valuable hypothesis. Students of sociology and political science in the course-of their studies would have come across analogies wherein society and state are compared to a biological organism, the natural law to the social law, thermodynamics to social dynamics, - etc. Such analogies, notwithstanding the fact, that as a class suffer from serious limitations, do provide certain insights which formulated as hypothesis stimulate and guide inquiries one of the recent orientations to hypothesis formulation is provided by cybernetics, the communication models, now so well entrenched in these social sciences testify to the importance of analogies as a source of fruitful hypothesis. The hypothesis that similar human types or activities may be found occupying the same territory was derived from plant ecology. When fine hypothesis was borne out by observations in society, the concept of segregation as it is called in plant ecology was admitted into sociology. It has now become an important idea in sociological theory. Such examples

may be multiplied. In analogy, may be very suggestive but care needs to be taken not to accept models from other disciplines without careful scrutiny of the concepts in terms of their applicability to the new frame of reference in which they are proposed to be employed.

(3) Hypothesis may rest also on the findings of other studies. The researcher on the basis of the findings of other studies may hypothesize that similar relationship between specified variables will hold good in the present study too. This is a common way of researchers who design their study with a view to replicating another study conducted in a different concrete context. It was said earlier that many a study in social sciences is exploratory in character, i.e. they start without explicit hypothesis, the findings of such studies may be formulated as hypothesis for more structured investigations directed at testing the hypothesis.

(4) A hypothesis may stem from a body of theory which may afford by way of logical deduction, the prediction that if certain conditions are present, certain results will follow. Theory represents what is known; logical deductions from this constitute the hypothesis which must be true if the theory was true. Dubin aptly remarks, "Hypothesis is the feature of the theoretical model closest to the 'things observable' that the theory is trying to model." Merton illustrates this function of theory with his customary facility. Basing his deductions on Durkheim's theoretic orientation, Merton shows how hypothesis may be derived as deductions from the theoretical system.

(1) Social cohesion provides psychic support to group members subjected to acute stresses and anxieties.

(2) Suicide rates are functions of unrelieved anxieties to which persons are subjected.

(3) Catholics have greater social cohesion than Protestants.

(4) Therefore, lower suicide rates should be expected among Catholics than among Protestants.

If theories purport to model the empirical world, then there must be a linkage between the two. This linkage is to be found in the hypothesis that mirror the propositions of the theoretical model.

It may thus appear that the points of departure vis-a-vis hypothesis construction are in two opposite directions (a) conclusions based on concrete or empirical observations lead through the process of induction to more abstract hypothesis and (b) the theoretical model through the process of logical deduction affects more concrete hypothesis. It may be well to bear in mind, however, that although these two approaches to hypothesis formulation seem diametrically opposed to each other, the two points of departure, i.e., empirical observations and the theoretical structure, represent the poles of a continuum and hypothesis lie somewhere in the middle of this continuum. Both these approaches to hypothesis-construction have proved their "worth. The Chicago School in American sociology represents a strong empirical orientation whereas the Mertonian and Parsonian approach is typified by a stress on theoretic models as initial bases for hypothesis-construction; hence hypothesis can be deductively derived.

(5) It is worthy to note that value-orientation of the culture in which a science develops may furnish many of its basic hypothesis. That certain hypothesis and not others capture the attention of scientists or occur to them in particular societies or cultures may well be attributed to the cultural emphases. Goode and Hatt contend that the American emphasis upon personal happiness has had considerable effect upon social science in that country. The phenomenon of personal happiness has been studied in great detail, if every branch of social science, the problem of personal happiness has come to occupy a position meriting central focus. Happiness has been correlated with income, education, occupation, social class, and so on. It is evident that the cultural emphasis on happiness has been productive of a very wide range of hypothesis for the American social science. Folk-wisdom prevalent in a culture may also serve as source of hypothesis.

The totality and substance of the discussion is aptly reflected in Larrabee's remark that the ideal source of fruitful and relevant hypothesis is a fusion of two elements: past experience and imagination in the disciplined mind of the scientist."

## CHARACTERISTICS OF A USABLE HYPOTHESIS

We now turn to address ourselves to the question 'what are the characteristics of a 'good' hypothesis?' Let us answer it; all hypothesis of science are seized by some kind of insight and intuition a leap of imagination. But the important methodological point is that a hypothesis so reached must be tested against data to be convinced of its adequacy. The power of a statement to explain its data determines its truth. But no amount of converging evidence can transform a factual statement into an analytical one and so free it from the principle of permanent control. The criteria for judging the usability of the hypothesis are none else than those that help the hypothesis perform their designated functions vis-a-vis research and the existing edifice of knowledge, generally. Hence a 'good' usable hypothesis would be the one which satisfies many of the following criteria:

- (a) A hypothesis should be empirically testable, it should be so stated that it is possible to deduce logically certain inferences close to the level of concrete observation so that they can be tested by observation in the field.

That is, the hypothesis should have empirical referents. The concepts embodied in the hypothesis must have clear empirical correspondence and should be explicitly defined. For example, 'Bad parents beget bad children' is hardly a statement that can qualify as a usable hypothesis, since 'bad' cannot be explicitly defined.

- (b) Hypothesis should be closest to things observable. Failing this, it would not be possible to test their accord with empirical facts. Cohen and Nagel rightly remark "...hypothesis must be formulated in such a manner that deductions can be made from it and consequently, a decision can be reached as to whether does or does not explain the facts considered."
- (c) The hypothesis must be conceptually clear. This point is implicit in the preceding criterion. The concepts utilized in the hypothesis should be clearly defined not only formally but also, if 'possible', operationally. Formal definition or explication of the concepts will clarify what a particular, concept stands for, while the

operational definition will leave no ambiguity about what would constitute the empirical evidence or indicator of the concept on the plane of reality. An ambiguous hypothesis characterized by undefined or ill-defined concepts cannot be tested since, understandably, there is no standard basis for knowing what observable facts would constitute its test. It is advisable that the concepts embodied in the hypothesis be redefined in a manner commonly accepted and communicable. This would ensure continuity in researches and go a long way in bringing about cumulative growth of scientific knowledge.

- (d) The hypothesis must be specific. One may hypothesize that something will happen in next five minutes, with absolute confidence but just because it is refuted it is empty of concrete information. We need to know what will happen and as soon as we commit ourselves to one view or another we become vulnerable; our prediction will be refuted if what we said would happen and as soon as we commit ourselves to one view or another we become vulnerable our prediction will be refuted if what we said would happen does not happen. A scientific statement is useful to the extent it allows itself to be exposed to a possible refutation. Often the researchers are tempted to express their hypothesis in terms so general and so grandiose in scope that they are simply not amenable to test.

This temptation can be suicidal. The researchers would do well to avoid employing concepts in their hypothesis for which suitable tangible indices have not developed. A hypothesis should include a clear statement of indices which are to be used. For example, the concept of social class needs to be explicated in terms of such indicators as income, occupation, education, etc. Such specific formulations have the obvious advantage of assuring that research will be practicable and significant. It also helps to increase the validity of the results because more specific the statement or prediction, smaller the probability that it will actually be borne out as a result of result of mere accident or chance.

Advisedly, the hypothesis should be related to a body of theory or some theoretical orientation. This requirement concerns the theoretic rationale of a hypothesis i.e., what will be the theoretical gains of testing the hypothesis. If the hypothesis is related to some theory, research will help to qualify, support, correct or refute the theory. A science can become cumulative only through interchange between the existing body of fact and theory.

Will not deriving hypothesis as a rule from some theoretical base throttle ventures into new fields in which no articulate theoretical system has developed? Will not such hypothesis lead to unnecessary repetitions? Doubts of this order may be raised by some. These objections do not have much substance since such hypothesis formulated imaginatively, besides serving the function of elaborating, extending and improving the theory, may also suggest important links between it and certain other theories. Thus, the exercise of deriving hypothesis from a body of theory may also be an occasion for a scientific leap into newer areas of knowledge. As Parsons put it, "Theory not only formulates what we know but also tells us what we want to know."

If hypothesis were derived from a body of theory, to that extent it would be possible to formulate them as statements about what will happen, that is, the roots of hypothesis in theory would invest these hypothesis with the power of prediction. One of the valuable attributes of a good hypothesis is its power of prediction. The potency of hypothesis in regard to predictive purposes constitutes a great advancement in scientific knowledge. To quote Cohen and Nagel, "... the hypothesis to be preferred is one which can predict what will happen, and from which we can infer what has already happened, even if we did not know (it had happened) when the hypothesis was formulated.

In the example cited earlier, the hypothesis that lower suicide rates should be expected among the Catholics than among the Protestants besides having a predictive potential power would also afford by virtue of its theoretical moorings, the basis for saying that married persons or a minority community or a tribal community by

virtue of high social cohesion would have lower suicide rates. It is in this sense that a 'good' hypothesis helps us make statements about what is already there or what has already happened although we were not aware of it.

(f) Hypothesis should be related to available techniques. This is, of course, a sensible methodological requirement applicable to any problem when one is judging its researchability. The researcher who does not know what techniques are available to test his hypothesis is in a poor way to formulate usable questions. In other words, the hypothesis should be formulated only after due thought has been given to the methods and techniques that can be used to measure the concepts or variables incorporated in the hypothesis.

This should not mean as implying, however, that formulation of hypothesis which are at a given time too complex to be handled by contemporary technique is a taboo at all times.

One must not forget that if the problem is significant enough as a possible frame of reference, it may be useful regardless of whether or not it is amenable to verification or test by the techniques-available at the time. The works of Marx and Durkheim have been of paramount importance to sociology even though at that time their larger ideas were not capable of being handled by available techniques. Lastly, it would be well to remember that posing of 'impossible' questions may stimulate the growth and innovations in technique. There is no doubt that some amount of impetus to modern developments in technique has come from criticisms-against significant studies which were at that time considered inadequate because of limitations of available techniques.

"How should the researcher ideally formulate hypothesis for research?" R.L. Ackoff has attempted to answer this question in a systematic way. As was stated in the previous chapter, all research problems ultimately reduce to the question, which of a set of alternative means is the most efficient one. Once these alternative means are formulated, the researcher is in a position to pose a question of each

of the means, as to what may constitute the evidence that this particular means is the most efficient one among the alternatives. The answer to this question would usually be in the form: "The particular means can be accepted as the most efficient among the alternatives under specific conditions." Such specific conditions should be formulated for each of the alternative means. The statements of these acceptance conditions are the hypothesis. The researcher does not of course, know which of these alternative hypothesis is true; this is precisely what the research is designed to determine. Ideally, a researcher should start with trying to determine all the alternative means (solutions or explanations) of coming to grip with his problem. This means that the researcher needs to undertake a, resource survey which includes survey of related theories orientations, which may bring to light what alternative means, solutions or explanations may be applied to the problem. As has already been said, the researcher will attempt to determine which of the alternative course of action or solutions and explanations is most efficient in terms of certain criteria e.g., economy predictability etc. Let us now suppose that a researcher has a problem whose solution depends on certain predictions and the researcher knows that there are three alternative theories (means) which are germane to the problem. Now, if one of the three theories is more likely to predict events accurately than the other two it may be taken as the most efficient one as a solution to the problem. If the problem happens to be one dealing with practical or programmatic concerns, the criterion of efficiency of alternative courses of action may be economy in the realm of time of time, money and energy. The alternative hypothesis which the researcher sets out to formulate are nothing but statements of conditions for each of the alternative means under which conditions, it (each alternative means) may be seen to most efficient. Simply stated, the alternative hypothesis are the statements of acceptance conditions for each of the alternative courses of action or alternative solutions to the problem. Suppose the researcher's problem is to decide which of the two types of teaching methods should be recommended for a particular educational institution. The research decides to use, let us say, the students'

examination scores as a measure of efficiency (of the means). Then, for each of the alternative teaching ' methods, his statement of acceptance conditions, i.e., alternative hypothesis, will be as under :

- H1: The average examination score produced by teaching method No. 1 is greater than the average test score produced by teaching method No. 2.
- H2: The average examination score produced by teaching method No. 2 is greater than that of teaching method No. 1. Therefore recommend No. 2 if H2 proves to be correct.

We note here that one possible outcome has not been considered, i.e., the test scores are equal for both teaching methods (No. 1 and No.2). Now, if the test scores were really equal, i.e., if both methods were equally efficient, the researcher will have no course of action to select for recommendation, consequently, he may have to add another course of action.

It is clear now that formulation of alternative hypothesis involves the following steps:

- (1) A measure of efficiency applicable to all the alternative courses of action is selected (Examination Score: sales, productivity etc.)
- (2) On the basis of this selected measure of efficiency a set of acceptance conditions for each alternative course of action is assigned.
- (3) The acceptance conditions are reformulated as hypothesis which are mutually exclusive and jointly exhaustive.

In all research (theoretical or action oriented) alternative courses of action (solutions, explanations) acceptance conditions (economy, predictions etc.) or hypothesis should be made explicit. In fact, if the acceptance of one set of alternative hypothesis rather than another would make no difference whatsoever to subsequent behaviour (scientific or practical) then the problem or its formulation is scientifically meaningless. It is obvious that there is no scientific way of selecting one of the alternative hypothesis as valid save

when there is some index of efficiency which can be applied to each of the alternative courses of action. The applicability of the measure of efficiency to alternative courses of action depends on certain conditions holding. For example, in our illustration of the alternative methods of teaching, the use of examination score as a measure of efficiency maybe suitable only if each student is allowed an equal period to complete the common test. Such conditions constitute the points of agreement among the hypothesis. These points of agreement among the hypothesis are either known or assumed to be valid. Should such an assumption be made, the researcher must make it explicit.

If the researcher sets up two hypothesis, there must be at least one point of agreement among them and one point of variance or disagreement.' These Iternative hypothesis may be represented symbolically as under :

H1 - MN1

H2 - MN2

H3 - MN3

H4 - MN4

The alternative hypothesis should encompass all possible outcomes of research, that is, they should be exhaustive with respect to the points of disagreement which will be tested. Secondly, of course, the hypothesis should be mutually exclusive. Failing these two requirements, research will not indicate which one course of action or solution should be selected from among the entire range of possibilities represented by the alternative hypothesis.

A very effective way of assuring ourselves that the hypothesis are mutually exclusive and jointly exhaustive of the universe of possibilities is to use the logical technique known as the "Boolean expansion." Suppose we have one common point of agreement (M) among the alternative hypothesis and three,points of disagreement (e.g., N,O and P), then the alternative hypothesis according to the requirements of exhaustiveness and mutual exclusion could be presented as under. The common point of agreement among these

hypothesis may be the examination score underspecific conditions. Thus, M = Examination score. The points of difference may be N = more than x; N' = less than x; similarly O = more than y, O' = less than y and P = more than z, P' = less than z.

(Read H4 as Examination score is more than x and y but less than z).

H1 - MNOP

H5 - MN'O'P

H2 - MN'OP

H6 - MN'OP

H3 - MNO'P

H7 - MNO'P'

H4 - MNOP'

H8 - MN'O'P'

In general, if there are n points of disagreements, there will be  $2n$  ( $2 \times 2 \times 2 \times \dots n$  times) alternative hypothesis in an exclusive classification. Only one of the can be true and must be true.

In a research involving more than two hypothesis, it is advisable to formulate the points of disagreement symbolically in a manner indicated to facilitate the construction of hypothesis. Intuition is often not a satisfactory guide.

Ideally there should be one hypothesis for each alternative course of action. Such a problem is one which involves estimation e.g., estimating an optimum number of workers for a production unit - 100, 250, 300, etc. The selection of the most efficient course of action depends on an estimate of the value of a critical variable (i.e., the exact' number of workers). In such cases, it is not economical to formulate explicitly each alternative course of action and to associate a hypothesis with each. We can use only a short-hand formulation. The alternate hypothesis can simply be stated as: "K. workers are needed" and the problem of research is to estimate this K. Now, since the estimate of the value of any variable is subject to error, it is advisable to express the estimate as a range of values rather than a single value, e.g., 300 + 50 workers 1250, 10 350) are needed.

One may now introduce the readers to the concept of null hypothesis. The alternative hypothesis and the null hypothesis together constitute the framework for the statistical testing of hypothesis. Suffice it here to say that the null hypothesis in its simplest form

asserts that there is no difference between two populations in respect of some property and that the difference found between the samples drawn from these population is only accidental and unimportant. H.M. Garrett remarks: "the null hypothesis is akin to the legal principle that a man is innocent until he is proved guilty. It constitutes a challenge and the function of a (research) is to give facts a chance to refute this challenge."<sup>1</sup> A null hypothesis in its other form may assert that the results found in research do not differ significantly from the results to be expected on a probability basis or as stipulated in terms of a certain theory. An alternative hypothesis may state, for instance, that:

H1 : The females commit suicide often than the males.

The null hypothesis ( $H_0$ ) may be stated as:

$H_0$  : The females and males do not differ in respect of the rate of suicide.

A null hypothesis is decidedly more useful than other hypothesis because it is exact. It is easier to disprove the contrary of an hypothesis than to prove it with complete certainty. In other words, while a hypothesis cannot be shown to be true without factual consequences, it may be falsified conclusively if the expected consequences fail to occur. This enables the researcher to eliminate some of the alternative hypothesis and while this is not finally establish the one he thinks is true, he can never be sure that he has identified and eliminated all the alternatives; still his confidence in will grow if it continues unrefuted. It might also be the case that it cannot possibly be refuted, but in this case it would not be of much practical use. according to Karl Popper, the real basis of science is the possibility of empirical dispro of. No amount of positive confirmation of the hypothesis, in Copper's view, entails its truth. One can only say that certain statements are definitely incorrect. Besides the statistical techniques are better adapted to testing a null hypothesis.

## **CONCEPTS-ITS ROLE IN SCIENTIFIC INVESTIGATION AND DEVELOPMENT OF SCIENCE**

The aim of science is comprehension as complete as possible of the connections between sense experiences in their totality and "the accomplishment of this aim by the use of a minimum of concepts...", A concept is an abstraction of observed things, events or phenomena. Mc Clelland defines the term concept as "a short-hand representation of a variety of facts. Its purpose is to simplify thinking by subsuming a number of events under one general heading. "Wherever any degree of abstraction is involved, whether in sophisticated scientific observation, in interpretative understanding or in the simple commonsense statements of fact, concepts are always used. Talcott Parsons rightly remarks: "..... there is no empirical knowledge which is not in some sense and to some degree conceptually formed.

Science is based upon the facts, i.e., such propositions as are supported by material evidence. Science not only observes them but is charged with the responsibility of stating them. For stating such aspects as constitute the subject-matter of a particular discipline, the science must have some terms. As suggested, the terms having a reference to things about which a science tries to make sense

are its concepts. Abraham Kaplan has put the matter succinctly. He says; "The important terms of any science are significant because of their semantics reach out to the world which gives the science its subject-matter. The meaning of such terms results from a process of conceptualization of the subject-matter. In this process, the things studied are classified and analyzed; several things are grouped together and particular things assigned to the several groups to which they belong. The concept 'Paranoid', for example, puts into a single class of certain set of persons and is itself analysed into such patterns as delusions of persecution, grouping together a set of actions, verbal or otherwise, as the case may be, and without regard to the actors performing them.

The language of science evolves in order to deal with the problems of its concern for which the ordinary language has proved inadequate. Each science thus develops and utilizes certain concepts in order that it may be able to pose a range of significant problems for investigation, conduct empirical observation, organize its data to provide answers to the problems and to communicate its findings. Of course, science employs as stated at the outset, the Occam's razor, the principle of parsimony, in its language. The rule of scientific analysis is, 'as few concepts as you dare and as many as you must'. The principle decrees that science should attempt to explain all its phenomena in the most economical way, employing as few terms, concepts, symbols or formulae as possible. The process of scientific investigation comprises, what to the investigator may appear to be flashes of insight, glimpses of conceptions between things or events and awareness of distinctions and differences. In order to be retained, these have to be symbolized as concepts if they do not initially occur in an already symbolized state. The ultimate goal of scientific investigations may be said to be the establishment of general theories which are indeed the conceptual schemes at a higher level of generality or abstraction.

To reach this level of theorizing or conceptualization, the investigator needs to conduct his empirical observations and to analyse these observations in a particular conceptual framework at the lower

levels. In sum, scientific enquiry involves understanding of the relationship among various concepts which symbolize objects, events and properties etc that constitute the subject-matter of a given discipline. Concepts are needed to formulate the problem and to design ways of solving it.

Some concepts are quite close to the objects or facts they represent. For example, the meaning of the concept 'Man' may be easily clarified by pointing to specific men. The concept is an abstraction of the characteristics of all men have in common; characteristics that are either directly observable or easily ascertained or measured. Other concepts cannot, however, be so easily related to the facts or phenomena to which they are intended or represent; motivation, need-disposition, attitude, role, adjustment, commitment, etc, widely used by social scientists are concepts of this kind. They are inferences, at a higher level of abstraction, from concrete events and their meaning cannot be easily conveyed by pointing to specific concrete objects, individuals or events. The greater the distance between the concepts used in a study and empirical facts to which they are intended to refer, greater the possibility of their being misunderstood and thus greater the care that must be exercised in defining them precisely. The social sciences have, admittedly fallen by the way at the greatest hurdle, viz., the exact definition of constructs.

A precise definition or explication of concepts being utilized by the researcher is a natural and sensible methodological requirement for any study. In one's own private world, it would hardly matter what one means by what. This personal idiosyncratic vocabulary which perhaps serves him alright as long as he is conversant with it, may keep things going. But in the world of the scientists, it would be disastrous if one's concepts did not mean the same thing to others as they meant to him. It would create serious problems of communication and correspondence among scientists and affect the cumulative and integrated growth of knowledge adversely, if the same concepts were used to describe really similar and indistinguishable entities or phenomena. This would create as one can easily visualize, confusion worst confounded. Such a confusion

does unfortunately exist in the larger domain of the social sciences which have borrowed a majority of the words from current popular terminology; even though the meanings attached to them have been sharpened or adapted to serve their more exacting use.

It should now be clear why the need for a clear-cut definition of concepts utilized by the scientist at various stages in his research has been laid so much stress on. Explication refers to the process of making explicit what is implicit in the concepts. In the absence of explication of concepts, others would not know what one is talking about, let alone whether he is talking about sense. Ackoff aptly states, "the function of a scientific definition of a concept is to make explicit the conditions under which and the operations by which we can answer questions about that which is conceptualized."

Scientific requirement calls for explication of concepts at two levels.

1. The concepts must be defined in abstract terms, giving the general meaning they are intended to convey. This may be called the 'formal definition'. Such a definition is designed to convey the general nature of the process or phenomenon in which the scientist is interested.
2. The concepts must be defined in terms of the operations by which they will be represented in a particular research. Such a definition is known as the 'operational definition'. Such definitions are designed to make possible the collection of data which the scientist or researcher is prepared to accept as the indicators or empirical referents of his concepts.

Let us discuss each of the two types of definitions in some details :

At first glance, providing unambiguous formal definition for the concepts might appear to be a very simple task. Such may be the case in respect of certain standardized concepts that are in scientific currency. But formal definition usually presents a host of difficulties of the kind Durkheim experienced. As John Madge points

out, Durkheim choose the topic 'suicide' as it could be easily defined. Although he started by asserting that defines 'suicide' fairly easy, almost at once, he experienced formidable difficulties. For example, if a man is so unhappy that he refuses to eat and wastes away, will this amount to suicide or an illness of some kind or, when soldiers march into battle in circumstances under which they are certain to be killed, would that be suicide? In formulating a definition, it is not always possible to be certain about whether to include particular instances. Durkheim after many years of examination of these different considerations, started with a preliminary definition. Subsequently, he found but its inadequate and kept on modifying it. His first definition was formulated thus; "the term suicide is applied to any death which is the direct or indirect result of a positive or negative act accomplished by the victim himself. But again, he realized that his definition neglected one important feature; namely, that suicide should be an intended act. He therefore, arrived at the following subsequent definition:

"Suicide is applied to all cases of death resulting directly or indirectly by a positive or negative act of the victim himself which he knows will produce this result. This is recognizable definition of suicide and augurs fairly well with the reckoning by which official suicide statistics are kept.

Let us now turn to a very simple case. Suppose we were interested in getting an answer to a 'simple' question in a housing survey. "How many people live in your house? A moment's consideration makes us to realize that we will be concerned with a variety of such questions. Should a son who is away at college most of the time during the year be included among those living in the house? Should a boarder who travels during the week and is in his room only on week-ends be included? Should the maid who sleeps in three or four nights and spends the others at her family-home be included? These questions amply testify to the difficulties one would come across in the course of providing formal definitions to concepts.

Lack of clarity in meanings of many psychological and

sociological concepts is considerable since, understandably, the nature of the language of science depends on its methods of observation. Says Theobald, "An investigator whose methods are limited to unaided perception will present his observation in the common description language which will not have critical precision. On the other hand, the indirect method of observation will require precise abstract language to present its case."

We shall here discuss the procedure required if adequate standardized formulation is not available. Comparing the definitions of a concept offered at different times in the history of science or different definitions offered at the same time, we generally find that there is a common core of meaning which runs through all the definitions. The meaning towards which various definitions of the concept point is the true meaning of the concept. Some scientists advocate looking for a definition with which most scientists would agree, but if this were to become a general rule there would be little progress in defining. We should, of course, seek to improve upon the definitions but improvement is possible only by taking the past and present definitions into account. The following procedure has proved very useful in arriving at better definitions of concepts.

- (i) Examine as many definitions of the concept, past and present, as possible. Keep chronology of the definitions in mind.
- (ii) Try to get at the core of the meaning toward which most of the definitions seem to point.
- (iii) Formulate a tentative definition based on the 'core'.
- (iv) If the tentative definition covers all the cases you think it should relative to your research objectives; where it does not, make necessary revisions.
- (v) Submit the definition to as wide a critical appraisal as possible. Include scientists as well as non-scientists from various fields among the critics.
- (vi) Make final revisions on the basis of the legitimate criticisms you receive.

It was stated earlier that formal definitions of concepts entering

the scientific corpus of a discipline are not enough. In addition, quite often the researcher-scientist needs to offer an 'operational-definition' to the concepts. To this point we now turn to the operational definition of a concept purports to specify the operations which observe, measure and record the phenomena symbolized by the concept.

Thus, the score on a particular scale for measuring I.Q. may be said to be the operational definition of the concept 'intelligence'. Samuel Stouffer, the Director of the 'American Soldier Studies.' defined the crucial concept 'personal adjustment' operationally, as follows :

".... It is assumed that, on the average, men who said they were in good spirits, that they were more useful in the army than the civilians, that they liked the Army, were better adjusted to the Army than men who were negative in several of these expressions."

Bradburn tried to define operationally, the elusive concept of Happiness, He states "one way to find out whether people are happy is to ask them." Respondents were accordingly asked to answer the following questions. 'Taking all things together, how would you say things are these days-would you say, you are very happy or not too happy'.

Now that the nature of operational definitions has been exemplified, the reader should be able to appreciate their value in scientific research. The operational definitions afford an understanding of the "true meaning" of the concept. "The true meaning", says Professor Bridgman, the inventor of operational definitions, "of the term is to be found by observing what a man does with it, not by what he says about it." What is important for us to know is what the researcher assumes to be the empirical indicators of his concepts. This is what really matters. No word has meaning if the circumstances under which it is used are not specified. A concept, on this view, is synonymous with the corresponding set of operations. What the researcher says social cohesion is, cannot be regarded as being of

much consequence; how he goes about assessing the degrees of social cohesion through the use of certain instruments is what constitutes the true meaning of social cohesion for purposes of his research. Thus operational definitions insist on meaning indeed and not in words.

Dodd, in the 'Dictionary of Social Science' states: "A definition is an operational definition to the extent that the definer (a) specifies the procedure including materials used for identifying or generating the definitions and (b) finds high reliability for consistency in applications of his definition."

Among sociologists for instance, operational definitions have been postulated by Chapin, Lundberg, Alpert and Kirkpatrick. In Alpert's words, no definitions have any meaning apart from the methods which have been employed. Says Lundberg, on the basis of such premises, space is that which is measured with a ruler, time is that which is measured by a stock, force is that which makes pointers move across the dial."

Very useful as these operations often are, we must constantly remind ourselves that an undue emphasis on such definitions is likely to defeat the ultimate purpose of social science research. Sociologists do not seem to be agreed among themselves about the importance of operational definitions. The critics have justly opined that the concept does not merely mean a set of operations. The operations are simply the techniques one has to use in order to get at and measure something behind these operations and the phenomenon itself. An operational definition is best applied to uncomplicated concepts such as time, temperature etc., but not to such complex ones as social cohesion, morale, love, etc. It can without doubt offer a high level of precision when applied to uncomplicated concepts. But it fails us just when we are in most need of clarity.

The crux of the matter with reference to operational definitions is that we do not say we are measuring unless we know what we are doing—that is unless we know both how to get the numbers and what the numbers stand for. A set of numbers, operationally generated,

means nothing unless it is taken as indication of a relationship between aspects of the world which in their theoretic setting are concepts.

It would, however be a wrong attitude to scale down operational definitions, for the social sciences depend on empirical confirmation which is best obtained by operational definitions.

After all our best efforts to reduce qualities into quantities through operationalization, the qualities essentially remain qualities and the magic of quantification does not change their innate texture.

Blummer rightly states that "the vagueness of a concept is equivalent to a difficulty in observing clearly a thing to which the concept is presumed to refer; indeed this difficulty of knowing what to observe, being able to observe it and knowing how to observe it, is the crucial obstacle in bringing the concept into touch with experience."

At this stage of social sciences, we find a compromise between these positions. We may deal with the problem more easily by remembering that the concept is a set of directions. It directs the communicant to a particular kind of experience, one which to some extent has been shared. If it does not do so, communication becomes hazardous.

## SPIRAL OF SCIENTIFIC METHOD

Commenting on the stages involved in Scientific Method, Ranganathan writes that in any subject or discipline, progressive steps of abstraction is itself an intellectual process in any discipline perpetually round and round in a never ending succession of cycles. He explained the process in four steps:

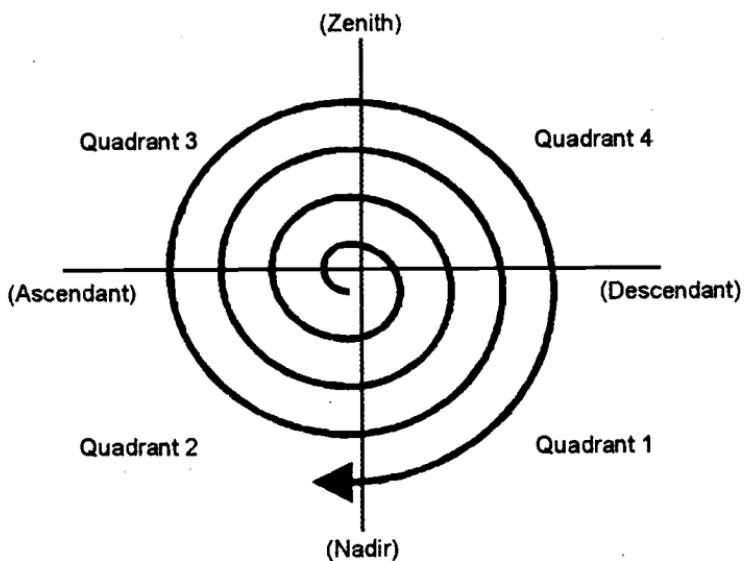
1. **Empirical Phase:** This phase starts with facts or individual experiences and ends in the distillation of few empirical laws. All this work is through intellect.
2. **Hypothesising Phase:** For this one must be endowed with a considerable intuition.
3. **Deductive Phase:** All the known empirical laws can be derived out of hypothesis by deductive reasoning and work in this phase is solely intellectual.
4. **Verification Phase:** The new deducted laws point to facts and experiments left unnoticed till then. The work in this phase consists of verifying them by observation in the phenomenal universe. It is done through the senses and intellect.

Dr. Ranganathan has characterized the working of the Scientific Method by a never-ending spiral movement. It is schematically represented in the diagram. The cycle implied in the spiral will be followed in the clockwise direction. For convenience of reference,

the four cardinal points of the cycle are denoted by Ranganathan by the terms Nadir; Ascendant, Zenith and Descendent.

The Nadir marks the accumulation of facts, obtained by observation, experimentation and other forms of experience.

The Ascendant marks the accumulation of inducted or empirical law got out of the facts accumulated at the nadir, by inductive logic including normal equations and other aids from statistical calculus.



**Fig. 8.1. Spiral of Scientific Method**

The Zenith marks the fundamental laws formulated with the aid of intuition of some degree or other so as to comprehend all the inducted or empirical laws accumulated at the ascendant as compelling implications.

The Descendent marks the accumulation of the descendent deducted laws got from fundamental laws at the zenith, with the aid of deductive logic including general emetics and all kinds of mathematical calculuses.

The four cardinal points give rise to four quadrants (a quarter of circle) in the cycle implied in the spiral. For convenience of

reference, the quadrants have been numbered 1 to 4 respectively by Ranganathan as follows:

**Quadrant 1 is the one between descendent and nadir.**

**Quadrant 2 is the one between Nadir and ascendant**

**Quadrant 3 is the one between ascendant and zenith, and**

**Quadrant 4 is the one between zenith and descendent.**

Ranganathan gives the description of the quadrants as under:

### **QUADRANT 1: EXPERIENCING FACTS**

Quadrant 1 corresponds to the stage in the development of a subject in the Universe of Subjects, conforming to Scientific Method, in which :

1. Primary senses are used either in their native state or with the aid of instruments of various intensity of powerfulness;
2. Observation are made of knowers, either with or without experimental interference and conditioning; and
3. Facts are found.

### **QUADRANT 2: INDUCTION AND GENERALISATION 1**

Quadrant 2 corresponds to the stage in which:

1. Intellect is used either by itself or aided by machinery constructed to speed up the work of the intellect and to give relief to it in some measure;
2. Reasoning is made with the aid of inductive logic including normal equations and statistical and other calculuses to boil down the numerous facts accumulated at the nadir to a small number of inducted or empirical laws for motion of planets, boiled down form the facts recorded by Tycho Brahe-more easy to hold in memory than the multitude of the facts themselves; and
3. Inducted or empirical laws are formulated.

### **QUADRANT 3: SUBLIMATION AND GENERALISATION 2**

Quadrant 3 corresponds to the stage in which:

1. Intuition of some intensity is used unmediated by the primary senses or the intellect;

2. The inducted or empirical laws stand boiled down to a very small number of fundamental laws—for example, Newton's Laws of Motion, to which many empirical laws such as the Kepler's Laws are reduced at one stroke—more easy to hold in memory than the more numerous empirical or inducted laws; and
3. Fundamental laws are seized.

#### **QUADRANT 4: DEDUCTION FROM FUNDAMENTAL LAWS**

Quadrant 4 corresponds to the stage in which:

1. Intellect is used either by itself or aided by machinery constructed to speed up the work of the intellect and give relief to it in some measure;
2. Reasoning is made with the aid of deductive logic including general semantics, and mathematical and other calculuses to work out all the compelling implications of the fundamental laws;
3. The deduced laws are derived;
4. The derived deduced laws include one and all of the inducted empirical laws; and
5. The number of deduced laws exceeds that of the empirical laws, if the fundamental laws had been seized with intuition of adequate intensity.

#### **ENTRANCE INTO THE NEXT CYCLE**

The Spiral of Scientific Method begins the next cycle thereafter, by re-entering in the quadrant.

Two things happen at the stage of development corresponding to quadrant 1 in the new cycle:

1. Observations and experiments are made to verify empirically the validity of the new deduced laws; and
2. Further observation and experiments are made continuously and this leads to accumulation of new facts.

## **DURATION OF VALIDITY OF FUNDAMENTAL LAWS**

So long as the deduced laws are empirically verified to be true and the new facts are found to be in conformity to the implications of the fundamental laws, there is no further movement in the Spiral, and the fundamental laws hold sway and continue to be deemed helpful. But this seldom holds good for long. Disturbance arises sooner or later in almost all the subjects in the Universe of Subjects as they get cultivated and developed. As and when new facts appear to contradict the fundamental laws, arduous effort is made to ensure absence of any kind of fallacy in the process of deduction, any kind of defect in the instruments of observation, more exacting experiments, and sharper calculuses for deduction, are devised. Witness, for example, of mass-spectroscopy which dissolved an apparent contradiction to Dalton's fundamental law about atomic weight, the devising of observations of star during total solar eclipse which removed an apparent contradiction of general semantics, and the mathematical calculuses to make deductions penetrating and free from fallacies. Every effort should also be made to reinterpret the terms in the formulation of the fundamental laws with the aid of exegetics, so as to get out of them their fullest possible implications, lying hidden until the necessary empirical experience draws attention to it.

## **DECLARATION OF CRISIS**

Now and again, contradiction between the facts and the currently used fundamental laws persists, in spite of all the possible precautions in the collection of the facts and the inference of the implications of the fundamental laws, and inspite of every possible and proper use of exegetics. Then we have to recognise the existence of a new class of facts and declare the incidence of a crisis in the application of the Scientific Method.

## **RESOLUTION OF CRISIS: NEW FUNDAMENTAL LAWS**

Then, the new class of facts are accumulated at the nadir; and a new cycle gets into full swing and carries the Spiral of Scientific Method further. Work in quadrant 2 leads to a new set of empirical laws. Work in quadrant 3 leads to the seizing of a new set of

fundamental laws. Work on quadrant 4 leads to a new set of deduced laws. Work in quadrant 1 is again started as earlier. This Cycle is liable to be repeated without end in many a subject.

## ILLUSTRATION OF RESOLUTION OF CRISIS

Newton's laws did not fit in with certain new facts found during the present century; and in this context, they ceased to be helpful hypothesis and gave place to Einstein's Law of Relativity. These two sets of laws belong to two different Cycles in the Spiral of Scientific method. In CC, the comprehension of all natural Sciences and their applications is comprehension of all Natural Sciences and their applications is scheduled as "A-Natural Sciences".

## HYPOTHESIS AND NORMATIVE PRINCIPLES

Any subject in the universe of subjects, irrespective of the idea or the method of study, admitting to the Spiral of Scientific Method in its development is a science. In the natural sciences, including the physical and the Biological Sciences, the fundamental laws are called "Hypothesis". They furnish a descriptive formulation of the facts of experience. In the Social Sciences, such as, Education, Political Science, Economics, Sociology, Law, and Library Science, the fundamental laws are called "Normative Principles". They furnish an interpretative explanation of the facts of experience. The terms 'true' and 'false' apply only to facts and the empirical and deduced laws. They do not apply to hypothesis and normative principles that is, to fundamental laws. These can only be described as helpful or unhelpful, according to as they fit in or not with the facts of experience.

### **Spiral of Scientific Method in Humanities**

#### **"SPIRITUAL EXPERIENCE". "FINE ARTS" AND "LITERATURE" NOT SCIENCES**

The Main Subjects "Spiritual Experience", and "Fine Arts" including "Literature", are products of intuitions with little play of intellect or sensory perception. Thus only Quadrant 3 occurs, The other three Quadrants do not come into action. So there is thus no Scientific Spiral at all.

### **“RELIGION” NOT SCIENCE**

In theology, the Zenith of the Spiral of Scientific Method is occupied by unalterable authoritative principles, often taken to be Revelations. Historically, even empirical experiences in the physical world, going contradictory to them, have been denied validity; indeed, history is full of instances of treating the assertion of such empirical experiences as heresay and of the cruelty meted out to heretics. Here in the left half the cycle—Quadrants 2 and 3 do not come into action. Exegetics is put to the severest strain to reconcile the authoritative principles at the nadir. One of the extreme devices used in this exegetics is known as Atharva Veda. According to this, the occurrence of meaningless words, phrases, sentences and even paragraphs are postulated to occur in a text. These are deemed to be unintelligible, and hence to be neglected as “noises” in the revealed texts occupying the zenith. Mimamsas in Indian Philosophy furnish outstanding examples of extreme forms of reconciliation of this kind. Christian theology also does so. There is thus, no Scientific Spiral at all. Therefore, it is not a Science.

### **“PHILOSOPHY” NOT SCIENCE**

In Philosophy, the Zenith of the Spiral of Scientific Method is occupied by intellectual constructs and not by intuitive apprehensions, though known as Hypothesis. Philosophy takes into consideration only deduced laws, facts of experience and empirical laws. If there is any crisis it is got over by changing the intellectual construct used as Hypothesis. Thus, Quadrant 3 does not come into action and there is thus no Scientific Spiral at all.

### **“LINGUISTICS”, “PSYCHOLOGY” AND “LITERARY CRITICISM”**

“Linguistics” and “Literary Criticism” have completed cycles of the Spiral of Scientific Method. Therefore, they are Sciences. But, they emanate from “Literature” and are auxiliaries to it. Therefore, it is necessary to keep them next to “Literature” which is not a Science. “Psychology” also has completed Cycles of the Spiral of Scientific Method. Therefore, it is also a Science. But, its historical origin was in Philosophy. Therefore, it is desirable to keep it next to Philosophy.

**EARLY RECOGNITION OF NATURAL SCIENCES**

Natural Sciences were admitted as sciences, as tested by the use of the Spiral of Scientific Method in their development, much earlier than Social Sciences. This was because, in Natural Sciences:

1. Observation-corresponding to quadrant 1 of the Cycle-can be impersonal and nearly objective, subject only to the errors of observation admitting of correction by the curve of errors and the calculus of observations designed more than a century ago by Gauss.
2. Experiments- corresponding to quadrant 1-can be made without any let or hindrance in the case of physical sciences and with a minimum of mental revulsion even with the living bodies of the lower organisms; and
3. The mathematical and statistical calculuses, and logic necessary and sufficient to make induction and deduction-needed in quadrants 2 and 4 in Natural Sciences, had been forged a few centuries earlier and there has been a continuous sharpening of them; and
4. Sublimation of the empirical laws into Hypothesis have been made in each of them;
5. New Hypotheses have also been formulated to resolve crisis.

**Applied Natural Sciences**

Some Applied Natural Sciences have completed at least one Cycle of the Spiral of Scientific method while the others have not yet done so. But an Applied Natural Science is a Pure Natural Science-in-action. Therefore, all Pure and Applied Natural Sciences should be kept in a group of their own.

**Social Sciences****DELAY IN THE RECOGNITION OF SOCIAL SCIENCES****DIFFICULTY 1: MAN HIMSELF BEING THE SUBJECT OF STUDY**

In the Social Sciences, unfavorable factors prevail in quadrants 1, 3, and 4.

1. In quadrant 1, observation is vitiated by the observed entity being man himself inducing obstruction in his mind to the pursuit of the subject;
2. Experimentation is difficult, if not impracticable, and even impossible when man himself has to be the subject of experiment; and
3. Even where it is attempted, it takes several decades to collect adequate data.

### **DIFFICULTY 2: PERIOD FOR EXPERIMENT**

The period for experiment with human beings has to be very long. For,

1. The life-span of a generation in human beings is several decades, as against the short one of rabbits (used as laboratory animals) and of still lower organisms; and
2. The rate of multiplication is small among human beings as against the fecundity of rabbits and phenomenally high fecundity of still lower organisms.

### **Difficulty 3: Humane Consideration**

Humane considerations prevent experimenting with the human species.

### **Difficulty 4: Delay in the Forging of the Necessary Statistical Aids**

Another difficulty has been the delayed emergence of statistical aids as shown below:

1. The statistical calculus necessary for work on humans, in quadrants 2 and 4 were either unknown or were found to be in a primitive state still about hundred years ago;
2. It was only in about 1890 that Statistical Calculus was developed and made fit for work, on humans, in quadrants 2 and 4 by Karl Pearson (1857-1936); and
3. Statistical Calculus, Operational Research, and other tools are being sharpened only today to a degree sufficient to make

allowances for the vitiations, incidental to man and his society being the subject of study.

## **SOCIAL SCIENCES QUALITY TO BE SCIENCES**

As a result of the forging of such powerful tools, the development of many disciplines in the Social Sciences is becoming progressively amenable to Scientific method. As a result of this, many Social Sciences have completed atleast one Cycle in the Spiral of Scientific Method. Here are some examples;

**Note:** For convenience of reference and for no other reason the CC numbers are used in the examples;

1. In the Schedules for “T-Education” and “X-Economics”, CC mentions different systems of developing the subjects. These give the successive sets of normative principles evolved in the formation of different cycles in the Spiral of Scientific Method;
2. In “Z Law” there are different systems of Jurisprudence with their corresponding Cycles in the Spiral of Scientific Method,
3. “W Political Science” has experienced three Cycles in the Spiral of Scientific Method with the successive formulation of the normative principles.
  1. “The King can make no mistake;” valid in a monarchy;
  2. “The Judiciary can make no mistake” valid in a Republic: and
  3. “The Legislature can make no mistake” postulated in legislative-action, about two decades ago, in South Africa, when the Judiciary struck down some of the enactments involving discrimination against colour people.
4. “2 Library Science” entered in the Spiral of Scientific Method on the formulation in 1928 and the publication in 1951 of the normative principles known as the Five Laws of Library Science.

## SOCIOLOGY

“Y-Sociology” has only reached the stage of empirical laws; in due course it may complete the first Cycle in the Spiral of Scientific Method.

## SOCIAL WORK

“YX-Special work” is an applied adjunct to “Y-Sociology”, Therefore, it should go with Sociology Consequently, it should also be included in the group of Social Science.

## SCATTERED SUBJECTS OF SOCIAL SCIENCES

In the other scattered subjects included in Social Sciences, the Spiral of Scientific Method has not yet started. They are too recent for its start.

The following table gives examples of grouping of subjects with the aid of the Spiral:

<i>SL. No.</i>	<i>Particulars</i>	<i>Examples of Subjects</i>
1.	Completed one or more Cycles in the Spiral (that is, all the four quadrants)	Physics, Chemistry, Biology, Psychology, Education, History, Political Science, Law
2.	Quadrants 1, 2, and 4 completed and the cycle likely to be completed in due course of time	Applied sciences such as Engineering, Chemical Technology, Animal Husbandry, Pharmacology, Sociology.
3.	Quadrants 3 and 4 alone	Pure Mathematics
4.	Quadrants 4 and 1 alone	Speculative disciplines such as Religion and Metaphysics.
5.	Quadrant 3 alone	Mysticism, Fine arts and Literature proper.

## CRITICAL EVALUATION OF SCIENTIFIC METHOD

The Scientific Method is a way by virtue of which one can test opinions, impressions or guesses by examining the available evidence both for and against them. It is the pursuit of truth evidence both for and against them. It is the pursuit of truth which is determined by logical considerations. Its basis is the bringing together of observation and hypothesis or of fact and idea. The process is cyclical, consisting alternately of improving the ways in which observations are made in revising the hypothesis. Observations are made in two ways: by making them in deliberately designed circumstances: i.e. as experiments, and by using apparatus to produce special required circumstances. Hypothesis are improved by making them simple, quantitative, and general. The process of successive refinement, both experimentally and theoretically, has no evident end. Scientific processes do not lead to absolute truth. New experience, additional data as evidence, constantly change them. The 'facts' of yesterday are outmoded today. They are replaced by generalisations that conform to the latest bodies of findings. In Physical Sciences such basic concepts as Space, Time, and Distance are constantly appearing with new connotation.

Evaluating the role of scientific method of research A. Neelameghan writes that it ensures systematic investigation of Subject. Scientific method makes possible an orderly process of investigation, the exercise of analytical thinking, and a critical approach to the study and development of a subject. It generally leads to publicly verifiable conclusions. The objective of the method is to examine and discover facts, without being vitiated by personal prejudices and hopes. Scientific method thus, provides a means of self-correction for a discipline at each stage of its development. In laying the foundation of a discipline and in developing it, the adoption of such a method in discipline is essential because it helps in reducing to a minimum, the number of faulty steps in an investigation. Moreover, the use of this method helps to develop the scientific attitude and the habit of systematic thinking in the investigator. Therefore, a discipline that is cultivated on the basis of the scientific method acquires a certain

amount of consistency and reliability in its structure and function as it develops.

A. Neelameghan has emphasized the helpfulness of the Spiral of Scientific Method on the mode of development of subjects as under:

1. It characterizes a discipline on the basis of its method of development rather than on the basis of its subject of study or of the method used in studying it.

The difficulties arising in the tradition method of characterizing a discipline as a science on the basis of the subject of study or on the method of studying it is by-passed;

2. Any discipline may adopt the scientific method in its development;
3. Certain specific attributes of the different subjects become clear. This is of help to the librarian in learning about certain affinity among large regions of knowledge such as those comprehended by each Basic Subject, Comprehensive Subjects, etc. This in turn, helps in placing these subjects in a helpful position in the Universe of Subject;
4. Different models of the development of subjects may be formulated to help the prediction of likely developments in each of the subjects or groups of them; and
5. The facility for the prediction of the trend of development in a subject is valuable to the classificationist and classifier in particular and the librarian in general.

In the development of a subject, the use of Scientific Method brings into play the sensory, intellectual, and the intuitive experiences of man. It is convenient to consider the method to take the form of a never ending spiral, consisting of experiment, observation, and collection of data, intellectual analysis of the sense data, abstraction of empirical principles with the aid of inductive logic, statistical methods and other modern techniques useful for the purpose, further generalisation to derive fundamental laws with the aid of deductive logic and semantics, further observations in the realms not previously

attempted, discovery of areas not in conformity to the fundamental laws, further researches followed by newer empirical laws which are in turn followed by more comprehensive fundamental laws, and so on. The Spiral of Scientific Method being a never ending one, the development of a subject is never complete, but ever continuing.

### **Limitations**

Scientific method is not free from limitations. Some of these are:

1. Scientific method involves abstraction.
2. Scientific method is never complete.
3. The conclusions arrived at by scientific method are not final.
4. Sciences have limited scope. Each science is concerned with a particular area and is based on certain assumptions.
5. Superstitions, cherished beliefs, etc., are hostile to the growth of scientific method. Institutionists, authoritarians, fictionalists and mystics often undermine the respect for scientific method.
6. Formal procedures are fruitless. Definitions and formal distinctions are not often used properly; and statistical information may be irrelevant and inconclusive.
7. Scientific judgement is difficult, and sometimes impossible, when situation demand immediate action.
8. The growth of scientific method in a society where there is no desire to know truth, or freedom for the expression of intellectual doubt, is surely hampered. "Fear of offending established dogmas has been an obstacle to the growth of astronomy and geology and other physical sciences."
9. The necessary time for reflection, and material for experiments are often lacking for the proper development of scientific method.
10. Scientific researches in social field are often in the hands of those who cannot always oppose the established opinions or taboos; and

11. No scientific method can guarantee certainty of achieving the goal and can prevent human life from being an adventure.

Not with standing its limitations, scientific method has its values and uses. Possession of knowledge is a universal human urge. But the desire to know is not often strong enough to sustain a critical type of enquiry. Institutions of knowledge cannot give any proof that opposition initiations are wrong. Scientific method is concerned with the verification of acquired knowledge. It finds out some order in which things are related together. The conclusion which is arrived at by the scientific methods has an objective validity. The objective nature of the scientific method is its greatest quality. Scientific method is the only way to increase the general body of tested knowledge and to eliminate arbitrary and ambiguous opinion. Scientific method springs from the desire to acquire truth, and when this desire is very strong in a community, the progress of scientific method becomes rapid and smooth. The method, however, may not always lead to the final destination of truth but, "enables large numbers of walk with surer step." It certainly minimizes the dangers associated with novelty, adventure and uncertainty. It lays down policies, and standard of moral judgement with a broader outlook than those of organic response or wild stimulus. It strengthens the love for truth and courage for overcoming illusions. It settles differences in a rational way, which is appealing to all. It is beyond a narrow outlook and subjective elements which are sometimes petty. It unites men through its noble and rational procedure. "Because it requires detachment, disinterestedness, it is the finest flower and test of a liberal civilization."

## **RESEARCH DESIGN**

A research design may consist of the following steps:

### **1. Title**

Indicate the title of the problem in precise terms.

### **2. Selection of the problem**

(a) **Source:** Indicate what suggested the problem to you.

(b) **Justification:** Justify the selection of problem by listing reasons

with respect to personal suitability and its general value.

- (c) **Scope:** Indicate the proposed scope of the study by stating briefly the kind or kinds of persons, situation, materials, and the like to which the study is to apply.

### 3. Procedure of Solution

- (a) **Logical analysis :** Given in systematic and in logically related arrangement the issue, elements, or sub-topics into which the problems may be divided.
- (b) **Research procedure:** Describe briefly, in related terms to the research procedures to be used, in attacking the several elements of the problem.
- (c) **Data needed:** List here the specific data, facts or information to be obtained.
- (d) **Procurement of Data:** Highlight how each part of data is to be procured including reference to the means by which personal contacts are to be made.
- (e) **Treatment of Data:** Show how the data are to be treated to be made meaningful as applied to the solution of the problem.
- (f) **Assumption made:** List the assumptions of the things taken for granted with respect to the several phases of the procedure employed in studying the problem.

### 4. Conclusions

- (a) **Hypothetical conclusions:** Indicate two or three of the conclusions or kinds of results that might grow out of the study.
- (b) **Implications of Conclusions:** Indicate some of the implications of the hypothetical results 'listed', i.e. give several inferences bearing on the problem of its probable applications.

### 5. Previous Studies:

Briefly summarize the previous related works done and list bibliographic references to research in the field or to related studies which might serve as models for the solution of the problem at hand.

Research design represents a compromise dictated by the many considerations that go into (social) research. It is not a highly specific plan to be followed without deviations, but rather a series of guideposts to keep one headed in the right direction. The research design in a way tells the researcher what observations to make, how to make them and how to analyse the quantitative representations of the observations.

## BASIC STATISTICAL METHODS, CONCEPTS AND TECHNIQUES

### STATISTICAL METHODS

Statistical method is comparatively of a more recent origin. It was first used in Biological sciences and Astrology but gradually it was put to use in social sciences too. At present statistical methods are widely used in Economics, Psychology and Anthropology. They are being increasingly used in the study of purely social phenomena, and attempts are being made to provide quantitative measurements to them. What is meant by statistical methods. Let us first look into the definition of statistical methods.

#### Definition

Statistics or statistical methods have been variously defined by different writers. In the year 1869, Mr. Quetlet prepared a list of 180 definitions given by various writers. It is neither essential nor useful to give all those definitions here. We shall, therefore, confine ourselves to some scientific and comprehensive definitions only.

1. Statistics is the science of measurement of social phenomena regarded as a whole in all its manifestations. -Bowley
2. Statistics is the branch of scientific method which deals with

data obtained by counting or measuring properties of population  
of natural phenomena.

-Kendall

3. The science of statistics is the method of judging collective natural or social phenomena from the result obtained by the analysis of enumeration or collection of estimates. -King
4. Statistics is the science which deals with the collection, classification and tabulation of numerical facts as the basis for explanation, description and comparison of phenomena.

-Lovitt

### **Characteristics of statistical method**

From the analysis of the definition of statistics we can deduce the following two fundamental characteristics of statistics.

#### **1. Aggregative Study**

Statistical methods are used for aggregative analysis and intensive study of individual unit is outside its scope. Statistical inferences are based upon averages and they are applicable upon the population as a whole and not on any particular unit. For example, the rate of increase in population is applicable to the population as a whole and not to individual families. In order to make statistical study, therefore, we must have fairly large number of items, without which the inferences would have no validity.

#### **2. Quantitative Study**

The other fundamental characteristic of statistical method that it is based upon quantitative analysis. The phenomena that is not capable of being expressed numerically can not be studied through statistical methods e.g. character, patriotism, attitudes etc. Attempts are now being made to provide standard quantitative measurements to the phenomena that have hitherto been mainly qualitative in nature. For example, standard of living intelligence and intensity of likes and dislikes have been measured quantitatively through various scales. This provides increasing possibility of the use of statistical methods in social research.

## Basis of Statistical analysis

Statistical analysis is based upon following two fundamental principles :

1. Law of statistical regularity
2. Law of inertia of large numbers

### 1. *Law of Statistical Regularity*

It is based upon the mathematical law of probability. According to this law a phenomena is influenced by a large number of causative factors. These factors are mutually independent and therefore, very seldom change in the same direction. Because of their pull in different direction their effect is to a great extent neutralised. We shall illustrate our point by a few examples. Suppose we toss a coin one thousand times, it is not possible that the coin may fall all the thousand times towards head or tail. Theoretically we can imagine such a situation but practically it is not possible. The probability is that the coin will fall nearly 500 times, towards head and an equal number of times towards tail.

Now we take another example from the social sciences. Suppose we have to draw a sample of one hundred students from a total of 10 thousand. It is theoretically imaginable that all of them may be rich or all of may be short statured. In practice, however, this will not happen. If the number has been randomly drawn the sample will contain both rich and poor, long and short student and in all probability in the same proportion in which they exist in the whole population of ten thousand students.

Law of statistical regularity makes the drawing of representative sample possible which is the basis of all statistical analysis. It also gives validity to statistical inferences. It tells us that what has been true of the past may be true of future also unless some fundamental change has taken place in the composition of variables. It is on the basis of this law that we try to project existing data into future.

### 2. *Law of inertia of large numbers*

The other fundamental law upon which the theory of statistical

analysis is based is the Law of Inertia of large numbers. According to Prof. Bowley - Large numbers and the average as we generally get from the measurement of social phenomena have great inertia. At another place he points out, "individual units of a group are constantly changing but the whole group changes very slowly." For example let us take the rate of growth of population or the ratio between males and females. In a single family the ratio between males and females may differ widely. Some of the families may have as many as six or seven children and all of them either males or females. But when we take the country as a whole the ratio between them is nearly equal. The rate of growth of the population of the country would be very much stable.

The greatest use of this law is that it makes the measurement possible. According to Prof. Bowley, it is this 'stability of large number that makes the measurement possible.' What is the use of collecting past data if it can not be used for some practical purpose. If the rate of growth of population were to change too often what is the use of collecting data about the rate of growth. But generally this does not happen. We can safely project past trends into future and make fairly accurate predictions. To take an example, there is nothing more uncertain than death. We can not predict when a particular person die, but we can quite accurately predict the death rate among a particular age group. An Insurance Company in U.S.A. once advertised, "We do not know who will die but we certainly know how many". This kind of accurate prediction is possible only through the operation of this law.

### **Functions of Statistics**

As a method of research, statistics performs the following functions :

#### **1. Providing measurement of phenomena**

Statistics provides measurement to social phenomena. In this respect it has two types of functions to perform. If there is already a scale of measurement we try to collect data according to it and if there is no standard scale of measurement we try to provide one

through statistical analysis and evaluation of variables involved. Thus the first category of functions includes collection of all types of data. Some of the data can be collected by means of actual counting while others have to be estimated e.g. the number of unemployed or invalids or beggars in the whole country.

The other function of statistics viz. providing standard scale of measurement where it does not exist is more important. Most of the social phenomena are qualitative in nature and we do not have standard scale of measurement. For example, we generally say that the standard of living of such person is high or low, but we can not give the exact measurement of it. Attempts are now being made to provide quantitative measurement to such social phenomena. Generally two types of methods are used in this respect : (1) index numbers and (2) scaling techniques. Index numbers provide a numerical measurement of the change in a composite phenomena overtime. Scales provide numerical measurement to quantitative facts.

## 2. *Description of facts*

Statistics provides description of fact by means of numbers. We can know about the magnitude of child marriage, or drinking through the statistics of these facts. Similarly we can have a clear picture of the unemployment situation in the country only when we have the figures of unemployed people, duration of unemployment, the type of work that they can do, size of their family, any supplementary source of income and so on. Statistics tries to introduce further clarity by means of the use of graph, diagrams, charts etc.

## 3. *Objective valuation of a phenomena*

Qualitative descriptions are generally subjective in nature and may differ according to researchers own idea of its magnitude. This gives rise to the lack of uniformity. Statistics, by providing standard scale helps eliminating element of subjectivity. Different people may give different impression regarding the crime situation in a country but when we express it in numbers there can be only one description. Statistics, thus helps in objective and accurate valuation of a social phenomena.

**4. *To ascertain trend and provide future estimates***

Statics tries to find out the direction and magnitude of change in a phenomena over time. With the help of these we can find out its position in the near or distant future by projecting the trend further. To take an example, we generally find that the population of a country tends to rise regularly. By measuring the rate of growth we can forecast population on any future date.

**5. *To provide comparative study***

Statistics provides the facility of comparative analysis. This comparison may be on the basis of time, place or facts. Comparison is made possible through quantitative measurement. For example the health of two towns can be compared through death rate. Intelligence of two or more students can be compared by means of I.Q. By giving the figures for the crime we can compare the administrative efficiency and police measurement helps in the comparison. The change in the price level over time can be compared by means of index numbers.

**6. *To ascertain and measure the degree of relationship***

With the help of statistical analysis we try to establish relationship between any two or more variables. This is done through various complicated statistical measures like coefficient of correlation, association of attributes, co-variance etc. These are highly technical methods and require previous knowledge. The more important thing about statistical inference is that we not only find out that two variables are correlated but we can also locate the degree of relationship.

**Process of Statistical Research**

The first stage in the statistical research is the collection of data. Apparently this seems to be a very simple task and so it is if the facts to be collected are few and can be easily counted. But at times we can have to collect figures spread over large areas, and actual counting is not possible. In such a case the figures have to be estimated. Statistics tells us how we can estimate these figures with reasonable accuracy.

Sometimes we have to deal with a phenomena that is qualitative in nature. Under these circumstances it may be necessary to find out some numerical measurement. For example, if we have to compare the standard of living of two or more persons we must first of all find out a numerical measurement of standard of living. When this has been done we can proceed to measure the standard of living of the people under study.

There are two methods of collecting data viz. (1) census method and (2) Sampling method. Under the census method the whole of the population is to be surveyed while in case of latter, a sample is to be drawn out. There are various methods of drawing the representative sample about which a detailed discussion has been given in the next pages.

After the data has been collected the next step is analysis or processing of data. The first step in this direction is classification and tabulation. Large figures may have to be approximated and rounded up to make them easy to handle. Classification may either on the basis of attributes e.g. division of population on the basis of sex, religion or race, or according to class interval e.g. division of boys in to different age groups.

After this the data may be subjected to a number of statistical computations in order to know more about the nature of data. Calculating average, dispersion, skewness etc. is the part of statistical analysis. All this is highly technical mathematical process which has been discussed in detail in the next chapters.

After the analysis is complete we try to compare and establish relationship between the different variables. Correlation, association of attributes and co-variance are some of the measures used for this purpose. Preparation of various kinds of index numbers, factor analysis etc. or other modes of analysis and comparison of data. Besides measuring and analysing existing data we also try to forecast their position on some future date.

The data has also to be displayed in form of graph, diagrams and tables. The use of maps, charts and pictures can also be made

for the purpose of display. The purpose of display is to make the comparison and only geometrically accurate, but also attractive to look.

### **Importance of Statistical method In Social Research**

The chief importance of statistical methods lies in providing the quantitative measurement to a phenomena. Lord Kelvin rightly says, "When you can measure what you are speaking about and express it in numbers, you know something about it, but when you can not measure it, when you can not express it in numbers your knowledge is of a meagre and unsatisfactory kind." Quantitative measurement is the sign of the growth of particular discipline and our knowledge and control over the phenomena. We are no longer satisfied by casual remark that the prices are rising, we must know how much they have risen. This we can do by means of index number of prices. It will no longer satisfy us to say that India has economically improved since independence. We would like to have an exact measurement. This can be provided by figures of national income and per-capita income. How can we know that police administration has improved, certainly by giving the figures regarding crime.

Quantification of social phenomena is the basis of objective observation. Qualitative descriptions is by nature haphazard, unstandardised and subjective. If two persons are asked to give about the standard of living of a person, they are very likely to give different options. This can be avoided only when we have found out an exact measurement of the standard of living. Similarly if intelligence of boys were to be expressed in the qualitative terms it would not give a clear picture to us. But if the same were to be expressed in terms of examination marks or I.Q. there will be no difficulty in understanding it and also there will not be variety of opinion about it.

Statistical analysis brings greater precision to our thinking. When facts are reduced to arithmetical figures all argument comes to an end and conclusion can be challenged only by counter statistics. Figures never lie. They will never distort anything. They will not over emphasize. They will put plain facts in the coldest and most

detached way whatever may be the outcome.

As we are moving more towards social planning we have to base our policy upon aggregative figures. This is not much of consequence whether a person has committed suicide under some strange circumstances, what is important for social planners is the fact whether there has been a fall in the number of suicides. We can never rid the crimes and their seriousness in increasing or decreasing. In spite of our best efforts the accidents must occur. We as social planners are mainly concerned whether the accidents have shown a declining tendency.

Statistics is equally important in the evaluation of social reforms and nature and of social evils. Nothing can give clearer picture of the evil of drinking than the figures regarding the cases of suicides, indebtedness, high death rate and incidence of disease in the families of those who are drunkards. Similarly the usefulness of prohibition can be judged by the facts.

Statistical methods are becoming more and more popular among the social sciences. Successful attempts have been made at providing standard quantitative measures of phenomena which has hitherto remained qualitative in nature. We are moving more towards perfection and precision. With the use of these refined tools of analysis.

### **Limitations of Statistical Methods**

Inspite of all merits of statistical method described earlier it has some serious limitations so far as its use in social sciences is concerned.

For the purpose of statistical analysis it is necessary that the phenomena should be capable of being expressed in numbers. It is no doubt correct that some progress has been made in this direction but still a major part of sociological problems lies outside its scope as it can not be quantified. It has not been possible yet to transform many of the subjective feelings into arithmetical numbers and it is hardly expected that it can be done in the near future.

MacIver has very strongly opposed the attempt at quantification

of social phenomena. His observations in this connection are worth noting. "It is common view that science begins and ends with measurement. Whether or not this is true of physical science, it is certainly not true of social sciences. The range of the measurable is not the range of the knowable. There are things we can measure like time, but yet our minds do not grasp their measuring. There are things we cannot measure, like happiness or pain and yet their meaning is perfectly clear to us. Perhaps after all we can measure only the external, the unknown and can know only the internal, the conscious state, the incommensurate. In which case sciences, if it is limited to the measurable, is limited to the measurable, is limited to the unknown perhaps even to the unknowable. It is then a means to power, not a understanding. It is only quantity we can experience. Science then counts the throbs of nature but does not feel them. They are to her as the beats of a vast mechanism, fulfilling eternally eternal laws. But the question, why ? - the only question that matters, the first question of childhood the first of age - that she can not answer."

Statistical method can at best tell us about the existence of some phenomena, it can not tell us why it exists in that form. We can know that the crimes have increased but we can not say why they have increased. Of course if we find that poverty has increased alongwith crime or there are more criminals among the poor we can establish some causal relationship between crime and poverty, but we can never say through statistical analysis alone as to which is the cause and which is the effect i.e. whether crime has encouraged poverty or poverty has encouraged crime.

Figures are too cold and feelingless to give a clear picture of society. We cannot understand society by any mechanical process of adding and subtracting. Thoughts and emotions of men have played a great part in changing our society but the same can not be put to statistical analysis. Statistics can touch only the fringe of the complex society. A far greater and vital part of it is still outside its scope.

## MEASURES OF TYPICAL SIZE-AVERAGES

Frequency distributions are extremely valuable for summarizing

large masses of data, but the process of summarization can be carried much further by compressing the characteristics of an entire series into one or at most a few significant figures. These figures are known as averages and they represent typical values of a variable. Averages hold a very important place in all types of statistical work. In fact, statistics has been referred to as "the science of averages." There are many different kinds of averages, but we shall consider only three, namely (1) the mean or arithmetic average (2) the median, and (3) the mode.

### 1. Arithmetic mean

Perhaps the most familiar of all the averages is the simple arithmetic average or mean. It is relatively easy to calculate and is widely used in statistical research. If the measure of each item in a series is known, the mean can be derived by adding the measures together and dividing by the number of items. If five students receive grades of 60, 75, 86, 88 and 96 respectively, on an examination, the mean grade is the quotient of the sum of the grades divided by five:  $405/5 = 81$ . The procedure in calculating the mean from ungrouped data can be expressed in algebraic terms by the following formula:

$$\bar{X} = \frac{\Sigma m}{N}$$

$\bar{X}$  is the mean.

$\Sigma$  (Greek capital s or sigma) is the conventional summation sign. It does not represent a separate quantity but indicates the "sum of" whatever follows.

$m$  is the measures or sizes of the separate items or cases.

$N$  is the total number of items or cases.

If there are relatively few cases, say less than 25 or 30, it is not very difficult to treat the items individually and compute the mean according to the above formula. If there are many more than 30 cases, and when other measures of central tendency and of dispersion are to be computed later, it is generally easier to arrange the data in a frequency distribution and calculate the mean in that way. There

are two common methods of computing the mean from a frequency distribution : (1) the long method, and (2) the assumed-mean or short-cut method. Since the assumed-mean, arbitrary origin, or short-cut method is so far superior to the long method and also since the same general method is used in calculating the standard deviation and the Pearsonian coefficient of correlation, it will be the only one considered in the present discussion.

**Table 9.1. Calculation of the mean ( $\bar{X}$ ) from a frequency distribution. Data represent weights of 265 male students at the Gurukul Kangri University.\***

<i>Class Interval (Weight)</i>	<i>f</i>	<i>d</i>	<i>fd</i>	
90-99	1	- 5	- 5	$\bar{X} = g + \frac{\sum fd}{N} \quad (i)$
100-109	1	- 4	- 4	
110-119	9	- 3	- 27	
120-129	30	- 2	- 60	
130-139	42	- 1	- 42	$= 145 + \left( \frac{99}{265} \right) (10)$
140-149	66	0	0	$= 145 + (.3736)(10)$
150-159	47	1	47	
160-169	39	2	78	
170-179	15	3	45	$= 145 + 3.74$
180-189	11	4	44	
190-199	1	5	5	
200-209	3	6	18	$= 148.74$
$N = 265$				
$\sum fd = 237 - 138$				
$= 99$				

\*Assumed data

The steps in the process of calculating the mean from a frequency distribution according to the short-cut method are as follows:

1. Arrange the data in a frequency distribution and ascertain the number of cases. From the illustrative problem in Table 9.1 it will be seen that the class-intervals, representing the weights of 265 Gurukul Kangri University students, comprise the first column and the frequencies the second column.
2. By mere inspection, estimate the interval in the distribution which is most likely to contain the mean. As far as the final results are concerned, any interval will do, but in order to keep the size of the numbers in the computation process to a minimum, care should be taken to select a class-interval as close as possible to the one that contains the mean, so far as one can judge. The midpoint of the assumed mean interval will be designated  $g$  in the formula. In the problem in Table 9.1 the interval 140 to 149 was assumed to contain the mean, and therefore  $g = 145$ .
3. In the third column the deviations ( $d$ ) from the assumed mean are marked off consecutively in steps or intervals. Those that are above the assumed mean in value are designated plus (+) and those below, minus (-).
4. The next operation is to multiply each frequency ( $f$ ) by its corresponding deviation ( $d$ ). The products are written in the fourth column ( $fd$ ). Of course, care should be taken to observe signs.
5. Find the algebraic sum ( $\Sigma$ ) of the  $fd$ 's. In Table 9.1  $\Sigma fd_{\text{neg}} = -138$  and  $\Sigma fd_{\text{pos}} = +237$ . Therefore, the algebraic sum of the figures in the  $fd$  column is  $\Sigma fd = +99$ .
6. Divide  $\Sigma fd$  by  $N$ . In the problem we have  $99/265 = .3736$  or, dropping the fourth decimal place as not significant, .374. The quotient obtained by this operation gives the correction in terms of the class-interval. Sometimes  $\Sigma fd/N$  will be positive and sometimes negative, depending on the position of the assumed mean.
7. In order to obtain the correction in terms of the actual units in the distribution, multiply  $\Sigma fd/N$  by the size of the class-interval

- (i). In the problem in Table 9.1, .374 is multiplied by 10. The result is 3.74.
8. The final step in the process is to add algebraically  $g$  and  $(N/\Sigma fd)$  (i). In the problem  $g = 145$  and  $(\Sigma fd/N)$  (i) = + 3.74, and the mean of the distribution, therefore, is 148.74 pounds.

## 2. Median

The median is another simple average or measure of central tendency. Many statisticians define the median as the *size* of the middle item when the items are arranged in their order of magnitude. This means, of course, that there are as many items above or larger than the middle one as there are below or smaller than it. If the number of items in the series is even, the median is taken as the arithmetic mean of the middle items. This concept which we are discussing should not be considered as a genuine median but rather as the *value* of the mid-item or mid-case. The concept "median" should be reserved for frequency distributions and not simple arrays. If such a distinction is made, then the median may be defined as that point on the scale of the variable (the point on the  $X$ -scale in a frequency graph) which divides the distribution into two equal parts.

In a frequency distribution the median is derived by interpolation in one of the classes of the distribution. There are two methods of interpolation : (1) the arithmetic and (2) the graphic. The arithmetic method is the one ordinarily used. However, the graphic method will be considered also since it will assist the beginner in grasping more clearly the significance of the concept.

First, summarize the steps that are followed in deriving the median ( $\bar{X}$ ) according to the arithmetic method.

1. The first and second columns in Table 9.2 are identical with those in Table 9.1.
2. The third column represents a cumulative frequency distribution. The frequencies in this column are accumulated from the top of the distribution. The cumulative frequency of any particular class-interval represents the sum of the frequencies of this and all the preceding class-intervals.

**Table 9.2. Calculation of the median (X). Data represent weights of 265 male students at the Gurukul Kangri University\***

<i>Class-Interval (Weight)</i>	<i>f</i>	<i>Cumulative f "Less than"</i>	
90-99	1	1	
100-109	1	2	
110-119	9	11	$X = l + \left( \frac{w}{f} \right) (i)$
120-129	30	41	$= 140 + \left( \frac{132.5 - 83}{66} \right) (10)$
130-139	42	83	$= 140 + \left( \frac{49.5}{66} \right) (10)$
140-149	66	149	
150-159	47	196	
160-169	39	235	$= 140 + (.750) (10)$
170-179	15	250	
180-189	11	261	$= 140 + 7.50$
190-199	1	262	
200-209	3	265	$= 147.5$
$N = 265$			
$\frac{N}{2} = \frac{265}{2} = 132.5$			

\* Assumed Data

- In order to determine the interval in which the median lies, the next step is to divide  $N$  by 2, ( $N/2$ ). In Table 9.2,  $N/2 = 265/2 = 132.5$ . It will be observed that 132.5 falls in the class-interval 140 to 149.
- Determine how many cases are required in this particular interval to reach  $N/2$  or 132.5. This is accomplished by subtracting the number of cases below the median interval

from  $N/2$ . By substituting the present illustration we have  $132.5 - 83 = 49.5$ . This number has been designated  $w$  in the formula.

5. Divide  $w$  by the number of cases ( $f$ ) in the median class-interval. In the illustration the frequency of the class-interval in which the median lies is 66. Therefore  $49.5/66 = .750$ .
6. Multiply this quotient by the size of the class-interval ( $i$ ). In the problem, ( $i$ ) is 10 and  $10 \times .750 = 7.50$ .
7. Add this product to the lower limit ( $l$ ) of the median interval. It will be observed that 140 is the lower limit of the median interval in the illustrative problem and  $140 + 7.50 = 147.5$ .

In order to derive a median by graphic interpolation it is necessary to construct either a "less than" or "more than" type (or both) of summation curve or ogive. The data for the "less than" type are accumulated from the top of the frequency distribution and the "more than" from the bottom of the distribution. The  $X$ -scale of the cumulative frequency graph is the same as that of the simple frequency graph, but the  $Y$ -scale is different in that the values must include a much wider range. In plotting the cumulative frequencies for the various classes the mid-points of the intervals are never used. For the "less than" type of curve the upper limits of the class-intervals are used and for the "more than" type, the lower limits. As will be seen from Figure 9.1 the typical smoothed ogive has the general characteristic of an elongated  $S$ .

To calculate the median for any distribution,  $N/2$  is first located on the vertical axis of the graph and an interpolation line parallel to the horizontal axis is drawn intersecting the curves. A perpendicular line is dropped from this intersection to the  $X$ -scale. The point where the perpendicular line cuts the  $X$ -scale is the median. It will be observed from Figure 9.1 that the two ogives intersect at a point which divides the distribution into two equal parts. This point, of course, is the same one through which the interpolation line is drawn.

### 3. Quartiles

Since the techniques for calculating quartiles are so very similar

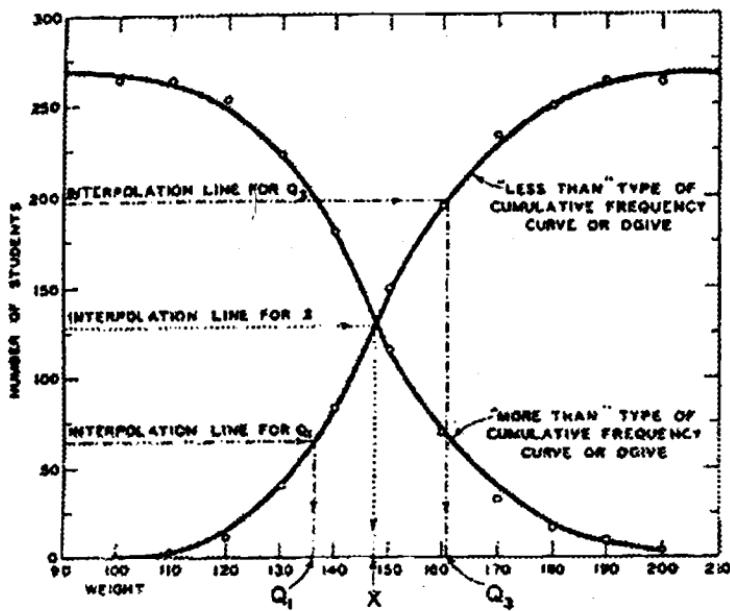


Fig. 9.1. Cumulative frequency curves or ogives or illustrating the graphic technique of interpolation for median and quartiles.

Based on data in table 9.1

to those used in deriving the median, there is much justification for discussing quartiles at this point. It should be understood, however, that quartiles are not measures of central tendency. Quartiles, deciles, percentiles, and other measures of this kind would more logically be included under the heading of variability.

It will be recalled that the median divides the frequency distribution into two equal parts. The quartiles, on the other hand, like the median, represent points on the scale of the variable. One of the three points is the median which actually can be designated  $Q_2$ , or the second quartile. The quartiles can be computed by either algebraic or graphic procedure.

The steps in calculating the quartiles according to the algebraic method are as follows :

- 1 The arrangement of the class-intervals, the simple frequencies, and the cumulative frequencies are the same as in Table 9.2.

2. In order to locate the intervals which contain the three quartiles,  $N$  is first divided by 4.  $N/4$  locates  $Q_1$  (the first quartile);  $(N/4)$  times (2) locates  $Q_2$  (the second quartile or median); and  $(N/4)$  times (3) locates  $Q_3$  (the third quartile).
3. After the class-intervals containing the quartiles have been located, the same formula used in interpolating for the median is applied. As an illustration let us compute  $Q_1$  for the data in Table 9.2. First,  $N/4 = 265/4 = 66.25$ . We observe, therefore, that  $Q_1$  lies in the interval 130 to 139. Substituting in the formula  $Q_1 = l + (w/f) (i)$  we have  $Q_1 = 130 + (25.25/42)(10) = 130 + 6.01 = 136.01$ .

Figure 1 illustrates the graphic method of calculating quartiles from summation curves. In order to derive  $Q_1$ , for example, the value  $N/4$  is plotted on the vertical scale of the chart and a horizontal interpolation line is drawn through the "less than" ogive. A perpendicular line is then dropped from the intersection of the interpolation line and the ogive. It will be seen from Figure 9.1 that  $Q_1$  is approximately 136.0.

#### 4. Mode

In a simple series the mode is the *size* of the measurement that occurs most frequently. In the following series of values : 15, 17, 18, 22, 24, 25, 25, 25, 27, 28, 28, 30, and 31, the mode is 25, because 25 occurs more often than any other measure. In a frequency distribution the mode is that point on the scale of the variable where the frequency is greatest. The beginner should never forget that in a frequency distribution the mean, median, and mode are values on the scale of the variable. In a frequency curve they will be represented by points on the X-scale.

**Table 9.3. Calculation of the mode ( $\bar{X}$ ). Data Represent  
weights of 265 freshman students at the  
Gurukul Kangri University\***

<i>Class-Interval (Weight)</i>	<i>f</i>	
		$\hat{X} = 1 + \left( \frac{f_2}{f_2 + f_1} \right) (i)$
90-99	1	
100-109	1	
110-119	9	$= 140 + \left( \frac{47}{47 + 42} \right) (10)$
120-129	30	
130-139	42	
140-149	66	
150-159	47	$= 140 + \left( \frac{47}{89} \right) (10)$
160-169	39	
170-179	15	
180-189	11	$= 140 + 5.3$
190-199	1	
200-209	3	$= 145.3$

\* Assumed Data

In practice it will be found that frequency distributions often have more than one mode. Distributions which have one mode are referred to as being *unimodal*. Those with two modes are described as *bimodal*, those with three, *trimodal* or more generally, *multimodal*.

There are several methods which can be used in deriving the mode in a frequency distribution, but none is entirely satisfactory or universally accepted. We shall mention only two methods in the present discussion. First, there is the method based on inspection, by which the modal class is ascertained and the midpoint of the interval is considered the mode. This mode is frequently referred to

as the *inspectional* or *crude mode*. The second method is based on interpolation. After the modal class on interpolation. After the modal class has been selected, the exact location of the mode within the class is determined by means of interpolation. This procedure is most reliable when it is applied to continuous variables in which the distribution is fairly normal. The formula is based on the assumption that the exact location of the mode within the modal class is determined by the frequencies of the two adjacent classes. If the respective frequencies of the adjacent classes are identical, then the mode is the midpoint of the modal class-interval. If the frequencies of the adjacent classes are different, then the mode is pulled in the direction of the larger frequency away from the midpoint of the modal class.

The steps in the derivation of the mode by this method are as follows :

1. Find the modal class by inspection.
2. The next step is to ascertain the lower limit of the modal class. In the problem in Table 9.3 it is 140.
3. Divide the frequency of the adjacent class above the modal class ( $f_2$ ) by the sum of the frequencies of both adjacent classes ( $f_2 + f_1$ ). In the problem,  $(f_2/f_2 + f_1) = (47/47 + 42) = (47/89) = .53$ .
4. Multiply the quotient thus derived (.53) by the size of the class-intervals ( $i$ ). In the problem, .53 is multiplied by 10 and the product is 5.3.
5. The last step is to add ( $l$ ) and the derived figure from  $(f_2/f_2 + f_1) (i)$ . Therefore, in the problem,  $\bar{X} = 140 + 5.3 = 145.3$  pounds.

## **APPLICATION AND CHARACTERISTIC FEATURES OF MEAN, MEDIAN, AND MODE**

What are the merits and the defects of the different averages and when should each be used ? A satisfactory average should be (1) definite and rigorously defined, (2) simple and concrete, (3) easily calculated, (4) readily comprehended, (5) susceptible of algebraic treatment and (6) suitable for the degree of precision of the measurement

of the variables. Judging in terms of these criteria, the mean is no doubt superior to either the median or the mode, but in the final analysis the nature of the data and the purpose at hand should really determine the particular average that is chosen. No average can be considered the *best* under all circumstances. Furthermore, too great dependence should not be placed on a single value. An adequate description of a frequency distribution ordinarily requires the computation of two or more averages as well as other types of statistical measurements.

Let us briefly summarize some of the more important characteristics of the mean, the median, and the mode.

### MEAN

1. The mean is the best known and most frequently used average. Oftentimes it may be advantageous to use the mean merely because it is so well understood.
2. The mean is affected by the value of every case in the series. As a result, undue weight sometimes may be given to extreme and erratic items. For example, the mean wage of a group may actually be misleading if there are extreme and a typical items at either end of the scale.
3. From an algebraic standpoint the mean is superior to either the median or the mode.

### MEDIAN

1. The median is not influenced by the size of extreme items. It is based on the values lying immediately on either side of it. The median can be used very effectively when extreme items are likely to have an unduly disproportionate influence on the mean.
2. The median is easily calculated.
3. The median is generally less reliable than the mean and is not so well adapted to algebraic manipulation.

### MODE

1. The mode, like the median, is a position average and is not

affected by extreme items. It is therefore useful in those cases in which it is desirable to eliminate the effects of extreme variations.

2. The mode is frequently very difficult to locate.
3. The mode possesses little significance unless there is a distinct central tendency and unless it is applied to a relatively large sample.
4. The mode is not susceptible to algebraic manipulation.

## VARIABILITY

Another important concept in statistics is variability. The mean, median, and mode give only one essential characteristic of a frequency distribution—its typical size or central tendency. To say, for example, that the mean or median value of the residential structures in a particular city is \$10,000 does not give a very adequate picture of property values. In addition, it is essential to know how the values vary above and below the mean or median. Some structures may be valued at only \$1,500, whereas others may be worth \$50,000 or more. Furthermore, it is possible for several distributions to have the same average yet be markedly different in variability. In some distributions the cases may cluster very closely around the average, and in others they may be widely scattered. It is, therefore, very important to determine the spread of the individual values on either side of their central tendency.

In the present discussion consideration will be given to the following measures of variability : (1) the range, (2) the mean deviation, (3) the standard deviation, and (4) the semi-interquartile range.

### 1. Range

The range of an ungrouped set of measures is merely the difference between the size of the largest and smallest items. For example, the range of the series, 10, 11, 13, 16, 17, and 19 is 9, for  $19 - 10 = 9$ . Sometimes the range is expressed in terms of the size of the extreme values, such as “a range of 10 to 19” in the foregoing illustration. The range is a very crude and unstable measure of

variability since it depends entirely on the two extreme items in the series and indicates virtually nothing about the general form or profile of the series. Moreover, it cannot be used very effectively in a frequency distribution since the exact range cannot ordinarily be determined. If only the total scatter of the items of a series is desired, or if the data are too scanty to warrant the computation of a more reliable measure of variability, the range may be used.

## 2. Mean deviation

The mean deviation, or average deviation as it is frequently called, is the mean of the sum of the deviations (irrespective of signs) from some measure of central tendency. The mean is usually taken as the standard although the median or mode is sometimes used. Care should always be taken to specify the particular average that is chosen for computing the mean deviation. To illustrate the calculation of the mean deviation let us consider the following simple series of values :

$m$ , series of values : 2, 3, 5, 7, 10, 11, 12, 13, 14, 16, 17

d, the deviation from the mean : 8, 7, 5, 3, 0, 1, 2, 3, 4, 6, 7

$$\bar{X}_m = \frac{110}{11} = 10$$

$$\bar{X}_d = \frac{46}{11} = 4.18$$

It will be observed that the mean of the series was first derived by adding the numbers together and dividing by 11. The deviations of each value from the mean of the series ( $\bar{X}_m = 10$ ) are indicated in the second line. The mean deviation is the sum of these deviations (disregarding signs) from the mean divided by the number of cases in the series, or  $46/11 = 4.18$ .

When the mean deviation is calculated from a frequency distribution the following procedure can be followed :

1. Compute the measure of central tendency that is to be used as a standard.

2. Indicate in a separate column the midpoint of each class-interval in the distribution.
3. In the next column tabulate the deviations ( $d$ ) from the standard that has been selected to the midpoints of the several class-intervals.
4. In the last column tabulate the products of the frequencies of each class-interval times the deviation ( $fd$ ).
5. Add the figures in the  $fd$  column, disregarding signs ( $\Sigma fd$ ).
6. Divide  $\Sigma fd$  by the number of cases in the distribution.

$$A.D. = \frac{\Sigma fd}{N}$$

In actual practice it will be found that the mean deviation is of little value and very seldom used in social research. The main justification, therefore, in devoting so much space to the mean deviation is to assist the student in obtaining a clearer understanding of the concept of variability and especially the standard deviation, which is much more widely used and is far superior in every way to the mean deviation.

### 3. Standard Deviation

The standard deviation, like the mean deviation, represents a mean of deviation items. It is different, however, from the mean deviation in that the deviations are squared before being summed, the sum of the squared deviations is divided by the total number of observations (cases) in the distribution, and the square root is extracted from this quotient. The standard deviation is always computed from the mean, whereas the mean deviation may be computed from the mean, the median, or sometimes the mode. Algebraically these steps may be represented by the following formula :

$$\sigma = \sqrt{\frac{\Sigma fd}{N}}$$

The standard deviation provides a more refined and statistically important measure of variability than the mean deviation.

**Table 9.4. Calculation of the standard deviation ( $\sigma$ ). Data represent weight of 265 male Students at the Gurukul Kangri University\***

<i>Class-Interval (Weight)</i>	<i>f</i>	<i>d</i>	<i>fd</i>	<i>fd</i> <sup>2</sup>	
90-99	1	-5	-5	25	$\sigma = \left( \sqrt{\frac{\sum fd^2}{N} - \left( \frac{\sum fd}{N} \right)^2} \right) (i)$
100-109	1	-4	-4	16	
110-119	9	-3	-27	81	
120-129	30	-2	-60	120	
130-139	42	-1	-42	42	
140-149	66	0	0	0	$= \left( \sqrt{\frac{931}{265} - \left( \frac{99}{265} \right)^2} \right) (10)$
150-159	47	1	7	47	
160-169	39	2	78	156	
170-179	15	3	45	135	
180-189	11	4	44	176	
190-199	1	5	5	25	$= (1.8367) (10)$
200-209	3	6	18	108	
$N = 265$		$\sum fd = 99$	$\sum fd^2 = 931$		$= 18.37 \text{ or } 18.4$

\* Assumed Data

The standard deviation squared is another useful statistical measure, called *variance*.

The standard deviation is symbolized by the abbreviation *S.D.*, but much more frequently by  $\sigma$ , the small Greek letter *s*.

In computing the standard deviation from a frequency distribution, the following procedure is the one most generally used. It is known as the short method.

- It will be observed from the illustrative problem in Table 11.4 that the first column contains the class-intervals and the second column, the frequencies. The total number of cases ( $N$ ) is 265.
- Select as an arbitrary origin the interval in which the mean is most likely to occur, and mark off the deviations in terms of intervals above (plus) and below (minus) the assumed mean. The interval 140 to 149 was chosen as the zero interval.
- Multiply the frequency of each class by its corresponding deviation ( $fd$ ) and enter the products in the fourth column. Find  $\Sigma fd$ . The procedure thus far is identical with that followed in computing the mean.
- In the fifth column tabulate the products of ( $fd$ ) ( $d$ ), that is, the  $fd^2$  values.
- Obtain  $\Sigma fd^2$ . It will be observed that there are never any minus quantities in the  $fd^2$  column. In the problem,  $\Sigma fd^2 = 931$ .
- The correction ( $c$ ), which is  $\Sigma fd/N$ , is squared. In the problem,

$$\frac{\Sigma fd}{N} = \frac{99}{265} = .374 \text{ and } (.374)^2 = .1399$$

It will be observed that  $c$  is expressed in terms of the class-interval and root in the original scale unit.

- The next step is to make the proper substitutions in the complete formula of the standard deviation.

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - c^2} \quad (i)$$

It will be observed that  $c^2$  is always subtracted. After the calculations have been made under the radical sign, the square root is extracted and the results multiplied by the size of the class-interval ( $i$ ). In the illustration in Table 9.4 the standard deviation is 18.4. In order to check the accuracy of the calculations a different guessed mean can be chosen and the problem recomputed accordingly.

#### 4. Semi-interquartile range

The semi-interquartile range or quartile deviation is one-half the distance from the third quartile ( $Q_3$ ) and the first quartile ( $Q_1$ ). The semi-interquartile range is commonly symbolized by  $Q$ . The formula for the semi-interquartile range is :

$$Q = \frac{Q_3 - Q_1}{2}$$

The semi-interquartile range for the distribution in Table 11.1, therefore, is :

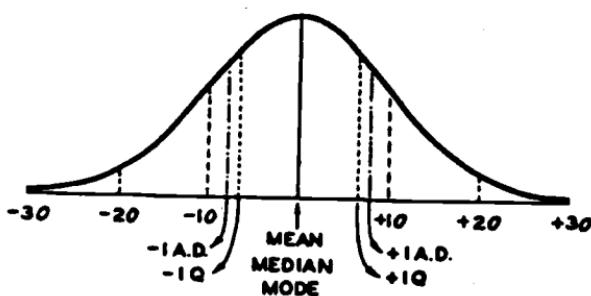
$$\begin{aligned} Q &= \frac{160.71 - 136.01}{2} \\ &= \frac{24.70}{2} \\ &= 12.35 \end{aligned}$$

Unlike the mean deviation and the standard deviation, the semi-interquartile range is not measured from any central standard. It will be recalled that the quartiles divide the distribution into four equal parts.  $Q_1$  is a point on the scale of the variable below which there are 25 per cent and above which there are 75 per cent of the cases, and  $Q_3$  is located so that 25 per cent of the cases are above it and 75 per cent below it. If the frequency distribution is perfectly symmetrical and bell-shaped,  $Q_1$  and  $Q_3$  are equidistant from  $\bar{X}$ . In an asymmetrical distribution this relationship does not hold. the semi-interquartile range is not as satisfactory a measure of variability as the standard deviation since only a portion of the distribution is taken into consideration and the remainder disregarded, and for certain types of distributions the quartiles become indeterminate and unrepresentative.

### RELATIONSHIP BETWEEN THE MEAN DEVIATION, STANDARD DEVIATION, AND SEMI-INTERQUARTILE RANGE

It will be recalled that the mean, median, and mode represent

points or values on the X-scale, and that the average deviation, standard deviation, and semi-interquartile range represent distances on the X-scale. In a normal frequency distribution, (1) what proportion of the area or the cases is included within the limits of the respective measures of variability ? and (2) what is the relationship between the average deviation, the standard deviation, and the semi-interquartile range ? In a perfectly normal distribution the A.D. ( $\bar{X}$  or  $\bar{X} \pm A.D.$ ) defines the limits of the middle 57.5 per cent of the cases; the  $\sigma(\bar{X} \pm \sigma)$  marks the range of the middle 68.26 per cent of the cases; and the  $Q(\bar{X} + Q)$  indicates the middle 50 percent of the items.



**Fig. 9.2. Normal curve showing the relationship of measures of variability**

It will be observed from Figure 2 that in a perfectly symmetrical distribution, or in a distribution that varies from this type only moderately, the mean deviation is approximately four-fifths of the standard deviation and the semi-interquartile range is about two-thirds of the standard deviation.

Table 9.5 presents more exactly the relationship between these three measures of variability.

As an illustration for interpreting a measure of variability let us take the standard deviation of the distribution in Table 9.4.

The standard deviation was found to be 18.4. In a perfectly normal distribution the standard deviation, when measured off above and below the mean, includes within this range approximately 68.26 per cent of the cases in the distribution. This fact is also approximately

**Table 9.5. Relationship between the mean deviation, standard deviation, and semi-interquartile range as found in normal frequency distributions**

<i>Measure</i>	<i>Relationship</i>
$\sigma$	$= 1.2533 AD$
$\sigma$	$= 1.4825 Q$
A.D.	$= .7979 \sigma$
A.D.	$= 1.1843 Q$
$Q$	$= .6745 \sigma$
$Q$	$= .8453 AD$

true for distributions which are almost normal.

In the problem one standard deviation below the mean (148.7) would be 130.3 and one standard deviation above would extend to 167.1 on the scale; and it could be said, therefore, that approximately two-thirds (68.26 per cent if the distribution were normal) of the students in this sample weigh between 130.3 pounds.

### COEFFICIENT OF VARIATION

The mean deviation, the standard deviation, and the semi-interquartile range represent measures of absolute variability. It is also frequently necessary to measure the relative variability of two or more frequency distributions. In Table 6 are listed the respective mean ages, standard deviations, and coefficients of variation of four groups of women who gave birth to one or more children in Roorkee during the five-year period 1996 to 2000.

Which group shows the relatively highest degree of variability and which the least? By merely examining the standard deviations it would be impossible to say. When, however, the standard deviations of the several distributions are related to their corresponding means, it is possible to determine the relative amount of variability of a number of frequency distributions. Karl Pearson has worked out a simple measure of relative variability which is generally known as the coefficient of variation.

**Table 9.6. Means, standard deviations, and coefficient of variation of the age distributions of four groups of mothers who gave birth to one or more children in the city of Roorkee 1996 to 2000\***

Classification	$\bar{X}$	$\sigma$	C.V.
Resident married	28.2	6.0	21.3
Non-resident married	29.5	6.0	20.3
Resident unmarried	23.4	5.8	24.8
Non-resident unmarried	21.7	3.7	17.1

\*Assumed data

$$C.V. \text{ or } V = \frac{\sigma}{\bar{X}} \left( \frac{100}{1} \right)$$

In Table 11.6 it will be seen that the non-resident unmarried mothers show the least relative variability (17.1) in age and the resident unmarried, the highest (24.8). The coefficient of variation for the problem in Table 11.4 is :

$$C.V. = \left( \frac{18.367}{148.736} \right) \left( \frac{100}{1} \right) = 12.3 \text{ per cent}$$

## SKEWNESS

In actual practice, frequency distributions are rarely symmetrical; rather they show varying degrees of asymmetry or skewness. In a perfectly symmetrical distribution the mean, median, and mode coincide, whereas this is not the case in a distribution that is asymmetrical or skewed. Figure 9.3 presents illustrations of curves that are noticeably skewed. The one on the left is skewed negatively and the one on the right is skewed positively. It will be observed that in the curve with negative skewness the mean is less than the mode, whereas in the curve with positive skewness the mean is greater than the mode. The mode is not affected by either the size or the number of extreme items in a frequency distribution, but the mean is.

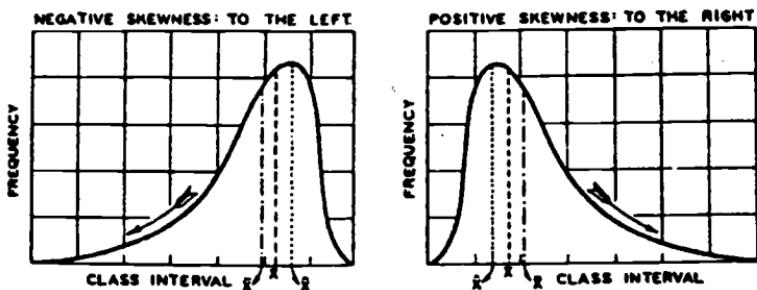


Fig. 9.3. Frequency curves showing negative and positive skewness

Coefficients of skewness have been devised which measure the direction of the skewness as well as the degree, either absolutely or relatively. A commonly used but only rough measure of skewness is merely the difference between the mean and the mode ( $\bar{X} - \hat{X}$ ). The sign indicates the direction of skewness and the difference, the amount. The distribution of the weights of 265 university students shows a slight positive skewness for  $\bar{X} = 148.74$  and  $\hat{X} = 145.28$  and  $148.74 - 145.28 = + 3.46$ .

A more satisfactory measure of skewness in which the degree of variability is given due weight is :

$$Sk = \frac{\bar{X} - \hat{X}}{\sigma}$$

Since the true mode is frequently difficult to determine, another formula developed by Karl Pearson for indicating skewness can be applied, especially if the series is only moderately skewed. According to this formula :

$$Sk = \frac{3(\bar{X} - \hat{X})}{\sigma}$$

The frequency distributions which have been considered thus far have been the common type in which the frequencies are relatively low at the ends and high toward the middle. In addition to this type

there are frequency curves which show a tendency constantly to increase or decrease. These curves possess the general characteristics of a capital "J" and are known as J-shaped curves. There is also another type of irregular distribution in which the frequencies are referred to as U-shaped curves. Both the J-shaped and the U-shaped types of frequency distribution do not occur nearly so often as the bell-shaped type.

## CORRELATION

In popular speech the idea of correlation is frequently expressed but generally in a qualitative and inarticulate manner. Comments are made concerning the relationship between criminality and feeble-mindedness, suicide and mental diseases, poor housing and morbidity, juvenile delinquency and the broken home, divorce and the business cycle, and between other phenomena. As students of social research we would naturally be interested in these problems, and should we attempt to study one of them, some type of statistical correlation probably would be utilized. In order to measure the relationship between only two variables, some type of simple correlation would be selected. For simple correlation a selection would be made from the following formulae depending upon the characteristics of the data and the problem at hand : (1) Pearsonian or product-moment, (2) rank-difference or rank-order, (3) contingency, and (4) curvilinear. If, on the other hand, the relationship among several variables was to be measured, partial or multiple correlation techniques or factor analysis would be applied.

The present discussion will be devoted largely to the Pearsonian or product-moment type of correlation, since it is the basis type of correlation and the one most commonly used. The product-moment coefficient of correlation ( $r$ ) is a pure number and ranges in value from positive one (+ 1.0) down through zero (0.0) to negative one (-1.0). That is, correlation may be direct or positive or it may be inverse or negative according to the direction of change, the size of the coefficient indicating the degree of relationship. When one variable increases (or decreases) and the other changes by constant or nearly constant amounts in the same direction, the relation of the two series

is positive; but if the changes in the two variables are in opposite directions, the correlation between the two series is negative. For example, the height and weight of human beings are positively correlated since taller people on the average weigh more than shorter people. On the other hand, in this country there is a negative or inverse relationship between socioeconomic status and fertility. Families with the larger income have relatively fewer children than families with smaller incomes.

It should be pointed out that the product-moment coefficient of correlation is based on assumptions that the data are derived from normal populations and that the association, if any, is linear. If there is considerable deviation from these assumptions, some other measure of association such as the Spearman rank-difference coefficient should be used.

In order to elucidate further the concept of correlation let us take the problem of measuring the relationship between the height and weight of college students. The data that will be analyzed have been derived from the freshman medical examination records of a sample of 265 men student at the Gurukul Kangri University.

1. The first step in computing a coefficient of correlation is to construct a scatter diagram. The scatter diagram portrays in graphic form the degree and type of relationship or covariation in the two series of data. Moreover, if the data are on a scatter diagram they can easily be transferred to the correlation chart for computation. In making a scatter diagram the first step is to select suitable class-intervals for the respective variables so that each will have from approximately 8 to 15 groupings. This procedure is similar to that followed in constructing a frequency distribution. In fact, the scatter diagram is a two-way frequency distribution.

From Figure 9.4 it will be observed that intervals of 10 have been chosen for the series representing weights and intervals of 1 for the series representing heights. The class-intervals for the X-variable read from left to right, and unlike the conventional frequency

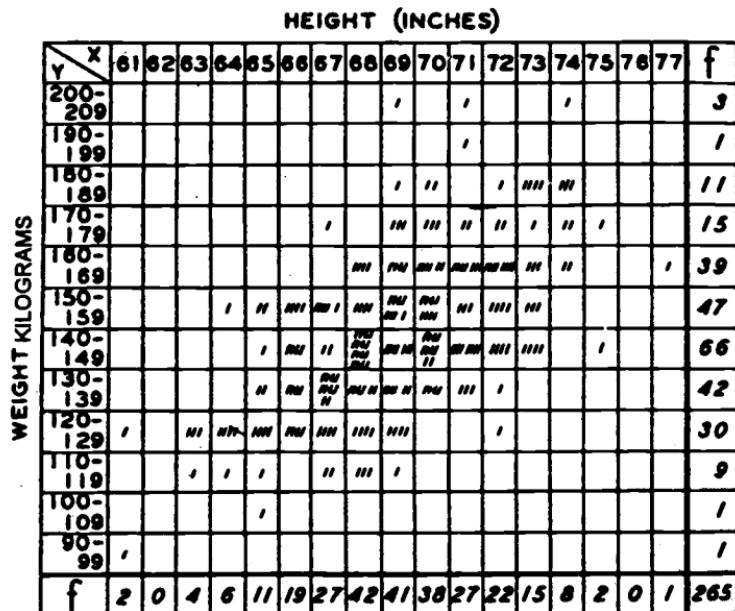


Fig. 9.4. Scatter diagram showing the relationship between heights and weights of 265 fresh men at Gurukul Kangri University

distribution, the intervals of the Y-variables read from bottom to top.

- Each entry in the field of the chart always represents two numerical values. In the illustrative problem one value will represent height and the other, weight. For example, the first case (student) chosen from our sample happens to be 66 inches in stature and 145 kilograms in weight. The value for height would of course be placed in the column representing the class-interval 66. The exact cell is located by selecting appropriate row, which would include the corresponding value for weight. For the above illustration a weight of 145 kilograms would be included in the row designated by the interval 140 to 149. The paired values for each case are plotted according to this procedure.
- It will be observed that the dots show a tendency to cluster in a wide band running from the lower left-hand corner to the

upper right-hand corner of the diagram, indicating that the correlation is positive. If, on the other hand, the cases were distributed along a band extending from the upper left-hand corner to the lower right-hand corner the correlation would be negative.

After the scatter diagram has been completed, and the distribution has been judged to be rectilinear, the data are then transferred to some standard product-moment correlation chart. It is much cheaper, quicker, and more reliable to use a printed correlation chart than it is to lay one out by hand. The correlation chart that will be used in the present discussion was devised by Professor F. Stuart Chapin, formerly of the University of Minnesota.\*

The following instructions summarize the various steps to be observed in computing a coefficient of correlation with this type of chart.

1. The class-intervals for both the X- and Y-variables should be written in the spaces at the top and the left-hand side of the correlation chart and the number of cases recorded in the proper cells. The frequencies for the two variables should also be entered on the chart. This operation involves merely a transferral of the essential data from the scatter diagram to the correlation chart. In selecting the zero-intervals for the two variables an attempt should be made to choose intervals in which the means of the respective distributions are most likely to occur. In this problem 69 was chosen to represent the zero-interval for the X-variable and 140 to 149 for the Y-variable.

2. For the X-variable determine the products of ( $f$ ) ( $d_x$ ) and record

---

\* All of the printed forms for computing the coefficient of correlation are very similar. A few of the better known charts are : (1) Thurstone (published by C.H. Stoelting Company, Chicago); (2) Otis (World Book Company, New York; (3) Cureton and Dunlop (Psychological Corporation, New York); (4) Tryon (University of California); (5) Ruch-Stoddard (University of Iowa); (6) Holzinger (University of California); (7) Dvorak (Longmans, Green, New York). (8) Kelley (World Book Company, New York); and (9) Durost-Walker (World Book Company)

them in the  $fd_x$  row. Multiply ( $f$ ) ( $d_v^*$ ) for the Y-variable and enter in the  $fd_v$  column. Care should be taken to observe signs. Determine the algebraic sums of the  $fd_x$  row and the  $fd_v$  column. In Figure 11.11 it will be observed that  $\Sigma fd_x = + 11$  and  $\Sigma fd_v = + 99$ .

3. The eruptive values of  $fdx^2$  are next obtained, as are also the values of  $fd_v^2$ , and recorded on the chart. Add the  $fd_x^2$  row and the  $fd_v^2$  column. In the problem,  $\Sigma fd_x^2 = 1,839$  and  $\Sigma fd_v^2 = 931$ . It will be recalled that the second and third steps are identical with those used in computing the standard deviation.
4. The fourth operation is different from anything that has thus far been discussed. First, note the small figures printed in the upper left-hand corner of each cell. Second, observe the signs for each of the quadrants indicated in the center of the field of the chart. The lower left-hand quadrant and the upper right-hand quadrant are plus (+) and the other two are minus (-). Multiply the number of cases in each cell by the corresponding printed figure in the cell, observing signs. The products are entered in either the  $\Sigma fd_{xy} +$  or the  $\Sigma fd_{xy} -$  column, depending on the sign. Let us illustrate this step by performing the computations in the row 130 to 139 in Figure 9.5. Multiply each of the frequencies in the row designated by the class-interval 130 to 139 by the small printed figures in each of the corresponding cells. The products for the numbers located in the plus quadrant are as follows : (4) (2) = 8; (3) (5) = 15; (2) (12) = 24; and (1) (7) = 7. The sum of these products, which is 54, is entered in the  $\Sigma fd_{xy} +$  column. The products of the numbers for this row that are located in the minus quadrant are : (3) (1) = -3; (2) (3) = -6; and (1) (5) = -5. By adding these products together we have -14, which is recorded in the  $\Sigma fd_{xy} -$  column. The figures in each of the columns are added and entered on the chart. The next step is to determine the algebraic sum of  $\Sigma fd_{xy} +$  and  $\Sigma fd_{xy} -$ . In the problem the figures are : 848 - 75 = 773.

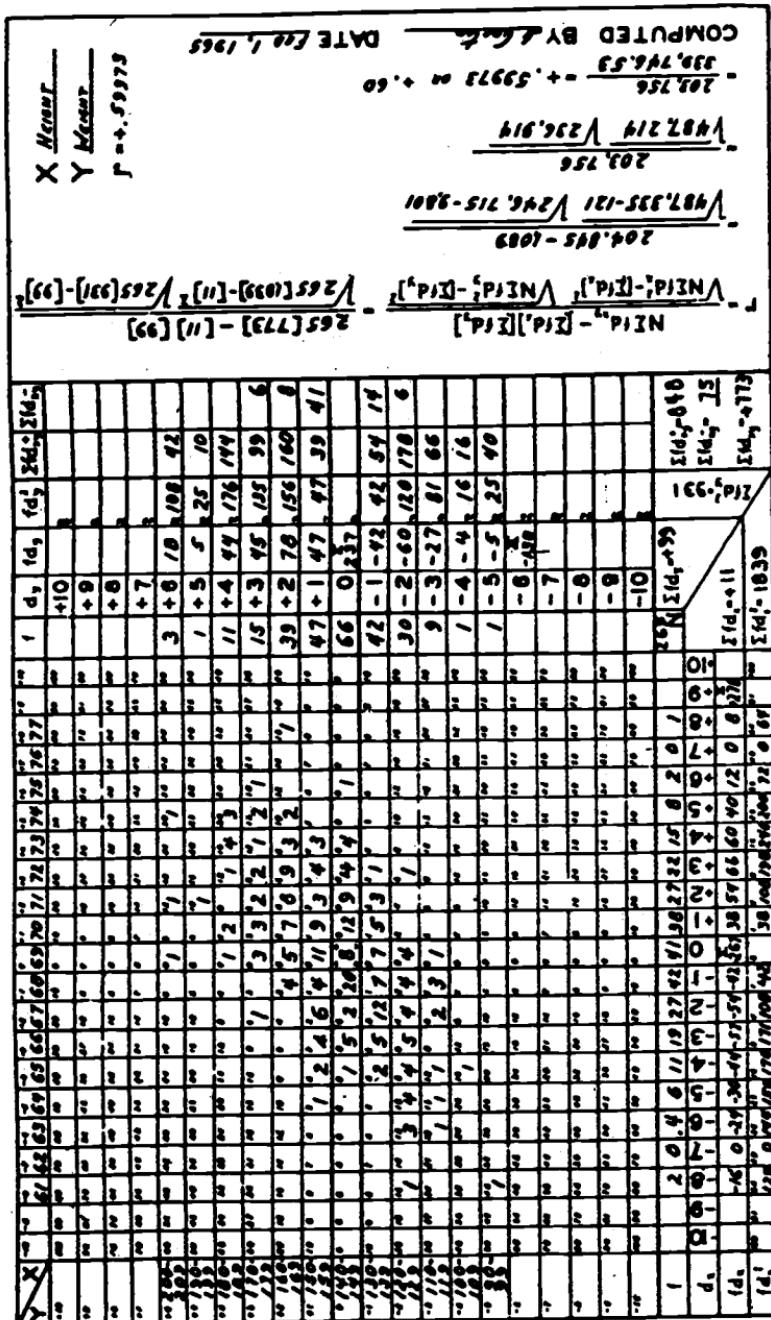


Fig. 9.5. Correlation chart

5. This completes all the preliminary computations on the chart. The final step is to substitute in the formula on the right side of the chart and proceed with the calculations. It will be observed from Figure 9.5 that the proper substitutions have been made in the formula and the coefficient of correlation has been computed for the illustrative problem. The coefficient of correlation between height and weight for this sample of 265 men students is  $r = + .60$ .

**Calculation of Pearsonian or product-moment coefficient of correlation ( $r$ ) from ungrouped data :** If there are no more than 100 or possibly 150 cases, it is generally easier to calculate  $r$  by the ungrouped method. This statement is made on the assumption that a good calculating machine is available.

As will be observed from Table 9.7, the procedure in calculating  $r$  by this method is very simple. The problem in Table 9.7 is to determine the relationship between cancer mortality and a cultural index for the United States Death Registration Area.

1. The first column contains the names of the states, the second column the respective cancer rates ( $X$ ), and the third column, the corresponding cultural index values ( $Y$ ).
2. The fourth column represents the  $XY$  products.
3. The fifth column contains the  $X^2$  values and the sixth column the  $Y^2$  values.

$$\begin{aligned}
 r &= \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}} \\
 &= \frac{47(142,016.3) - (3586.6)(1740)}{\sqrt{[47(283,886.98) - (12,863,699.56)]}} \\
 &\quad [47(83,664) - (3,027,600)] \\
 &= \frac{6,674,766.1 - 6,240,684}{\sqrt{(13,342,688.06 - 12,863,699.56)}} \\
 &\quad (3,932,208 - 3,027,600)
 \end{aligned}$$

**Table 9.7. Calculation of Product-moment coefficient of correlation ( $r$ ) by ungrouped method. The relationship between mortality from cancer and other malignant Tumors and state Quartile distribution in 152 cultural items.\***

<i>State</i>	<i>Cancer Rate X</i>	<i>Cultural Index Y</i>	<i>Rate Times Index XY</i>	<i>Rate Squared X<sup>2</sup></i>	<i>Index Squared Y<sup>2</sup></i>
Aurnachal Pradesh	60.8	13	790.4	3,696.64	169
Delhi	65.0	41	2,665.0	4,225.00	1,681
Goa	46.9	10	469.0	2,199.61	100
Himachal Pradesh	91.1	93	8,472.3	8,299.21	8,649
Meghalaya	80.1	34	2,723.4	6,416.01	1,156
Mizoram	96.1	72	6,919.2	9,235.21	5,184
(In order to conserve space detailed data for only 10 states are included in this table.)					
Orissa	83.1	54	4,487.4	6,905.61	2,916
Tripura	63.0	19	1,197.0	3,969.00	361
Uttaranchal	89.6	38	3,404.8	8,028.16	1,444
Uttar Pradesh	62.9	50	3,145.0	3,956.41	2,500
Total ( $\Sigma$ )	3,586.6	1,740	142,016.3	283,886.98	83,664

\*Assumed data

$$= \frac{434,082.1}{\sqrt{433,296,829,008}}$$

$$= \frac{434,082.1}{658,252.8}$$

$$= .659445884 = .66$$

4. The sums of the figures in each of the columns are next determined.
  5. Proper substitutions are made in the formula and the coefficient of correlation computed. In the illustrative problem,  $r = +.66$ .
- Interpretation of the coefficient of correlation :** In interpreting

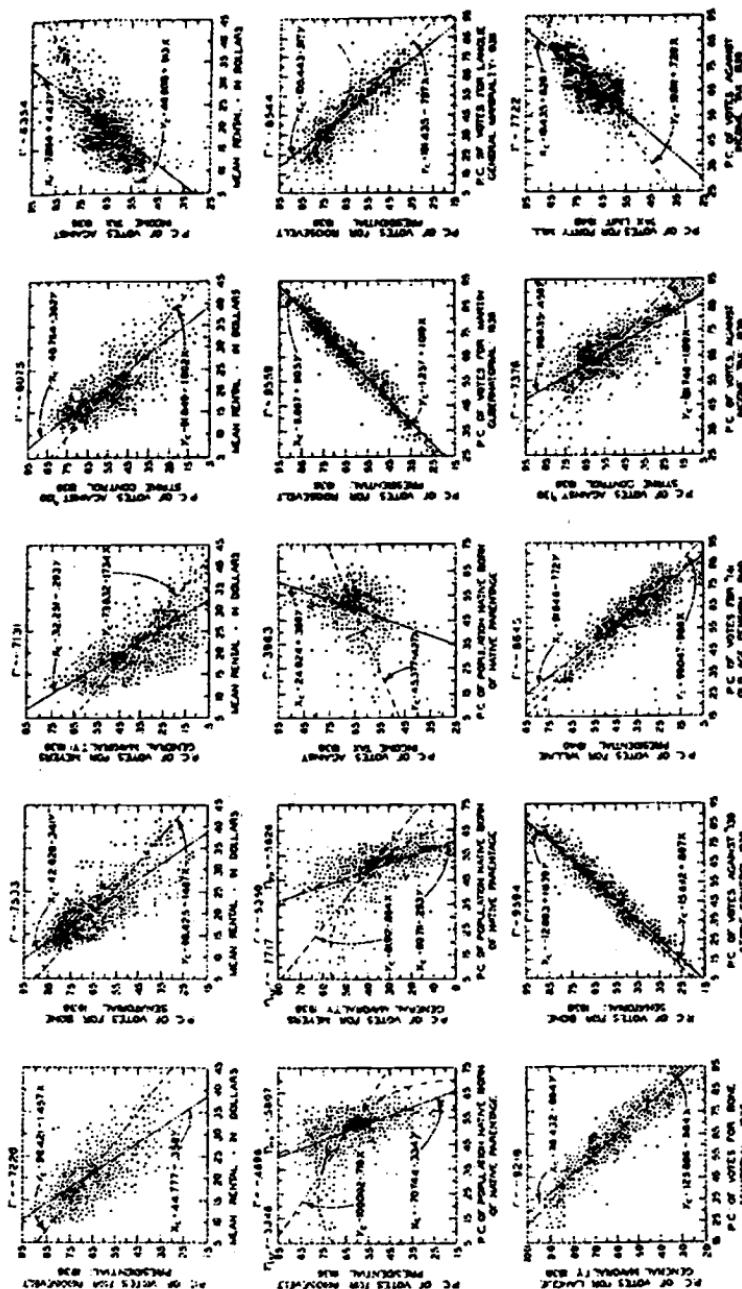


Fig. 9.6. Illustration of correlation techniques in sociological research. Note especially (a), (b) regression lines and equations, (c) rectilinear coefficients of correlation, and (d) curvilinear coefficients of correlation. Taken from Calvin F. Schmid, Social Trends Seattle, p. 274.

a coefficient of correlation the size of the coefficient is not the only consideration. Just as important as the absolute size of the coefficient are the relative size of the sample and the nature of the data. To say that a coefficient of correlation of .75 (plus or minus) is "high" may turn out to be very misleading if the standard error or probable error is relatively large. Or again, to say that a coefficient of .40 is always to be considered "low" may be erroneous for certain types of data. In judging the size of a particular coefficient, consideration should be given to other coefficients that have been derived from the same kind of data. It might be found, for example, that the coefficients of correlation for certain variables have been consistently less than .20 so that a coefficient of .40 would be considered relatively high for this particular type of data.

However, in order that the reader may have a few more definite facts which can be used for interpreting the size of the coefficient of correlation, if the sample large with 100 or more cases, the following general and tentative classification might be found to be of some assistance : (1) a coefficient of .70 to 1.00 (plus or minus) signifies that there is a *high* degree of association between the series; (2) if the coefficient is greater than .40 but less than .70, there is a *substantial* relationship; (3) if the coefficient is greater than .20 but less than .40, there is a *low* correlation, and (4) if the coefficient is less than .20, there is a *negligible* relationship.

The fact that two variables show a "high" correlation is no evidence *per se* of a causal relationship. A coefficient of correlation shows only the degree of association between two sets of phenomena. Whether or not the two phenomena are causally related is a matter of interpretation. There are at least three possible inferences which might be drawn in a case where two variables evidence a pronounced degree of correlation : (1) one might be the "cause" of the other; (2) they both might be related by one or more common "causes"; and (3) the correlation may have occurred by mere chance.

**Rank-difference method of correlation :** The rank-difference method of correlation may be found useful for relatively small samples -usually less than 30 cases-or where the ranks of the item of the two

variables are the only information available. From a mathematical standpoint the rank-difference method is decidedly less satisfactory than the product-moment method. For this reason the product-moment method should generally be preferred to the rank-difference method whenever the assumptions underlying the product-moment method are warranted. Although less precise and sensitive, the rank-difference method is independent of the distributions of normality and linearity which are imposed on the product-moment method.

In describing the procedure for computing a coefficient of correlation ( $\rho$ , Greek letter rho) by the rank-difference method, the problem in Table 9.8 may be considered by way of illustration.

1. The problem in Table 9.8 is to determine the relationship between the incidence of suicide and mobility of population in 25 large cities of the India. In the first column are listed 25 cities and in the second and third columns, respectively, suicide rates, and a mobility index.
2. The relative rankings of the various cities for each of the variables are recorded in the fourth and fifth columns. Ranking merely involves the numbering of the variate values according to the positions they occupy when arranged in order of magnitude. The highest variate value is given the rank of 1, the next highest, 2, and so on. In case of ties in rank, one of two methods can be followed, the "bracket-rank" method or the "mid-rank" method. In the "bracket-rank" method the items with the same value are assigned the same rank, and the next item after the ties is given the rank it would have had in case there had been no ties. In the "mid-rank" method the tied items are also given the same ranking, but the ranking represents the mean rank of the tied items. The latter method is generally preferable and is the one followed in Table 9.8. It will be observed that for suicide, New York and Milwaukee have a rank of 13.5, Cleveland has a rank of 12, and Baltimore, 15. Since both New York and Milwaukee have rates of 19.3 their rank is merely  $(13 + 14)/2$ , or 13.5. By the "bracket-rank" method both New York and Milwaukee would be given the

rank of 13, and the rankings of Cleveland and Baltimore would still be the same as in the "mid-rank" method.

3. The next step is to find the differences in the rankings ( $R_x - R_y$ ) for each of the items.
4. The differences in the rankings are next squared and the figures are recorded in the seventh column.
5. The sum of the seventh column ( $(R_x - R_y)^2$ ) is obtained and the proper substitutions are made in the formula in Table 9.8. It will be observed that  $\rho = + .54$ .

The rank-difference coefficient of correlation ( $\rho$ ) and the product-moment coefficient of correlation ( $r$ ) not identical, but they are similar in value. For example, in the illustrative problem  $\rho = + .54$ ; the approximate equivalent of  $r = + .558$ . Pearson has devised a correction formula which can be used to translate  $\rho$  into  $r$  or vice versa.

**Association between qualitatively defined variables :** It is frequently desirable in the social sciences to study the degree of association between attributes, or qualitatively defined variables. Although the problem of association is basically that of correlation, it does not seem likely that any measure of association between attributes will lend itself to mathematical manipulation and precision such as has been found in quantitative measures of correlation. Measures of association between attributes are limited to comparisons of predictability of a variable when its association with another variable is taken into account, as compared with its predictability when no association is utilized. A coefficient that performs this function may be considered to be a qualitative correlation coefficient, but more precisely should be called a coefficient of relative predictability.

**Table 9.8. Calculation of the rank-difference coefficient of correlation ( $\rho$ ). Relationship between suicide and mobility of population for twenty-five large Indian Cities<sup>\*</sup>**

City	Variables		Ranks		$R_x - R_y$	$(R_x - R_y)^2$
	Suicide Rate	Mobility Index	Suicide	Mobility		
	X	Y	$R_x$	$R_y$		
Agra	1.3	54.3	13.5	11	+ 2.5	6.25
Delhi	17.0	51.5	19	8	+ 11	121.00
Mumbai	17.5	64.6	16	17	- 1	1.00
Jaipur	16.5	42.5	21	5	+ 16	256.00
Ajmer	23.8	20.3	7.5	1	+ 6.5	42.25
Chennai	20.1	52.2	12	10	+ 2	4.00
Ahemdabad	24.8	62.4	4	16	- 12	144.00
Baroda	18.0	72.0	15	23	- 8	64.00
Meerut	14.8	59.4	23	14	+ 9	81.00
Muzaffarnagar	14.9	70.0	22	22	0	0.00
Dehradun	40.0	43.8	1	6	- 5	25.00
Haridwar	19.3	66.2	13.5	19	- 5.5	30.25
Najibabad	13.8	67.6	24	20	+ 4	16.00
Nagina	22.5	37.1	9	4	+ 5	25.00
Jaipur	23.8	56.3	7.5	- 12	+ 4.5	20.25
Ajmer	17.2	82.9	17.5	25	+ 7.5	56.25
Udaipur	23.9	62.2	6	15	- 9	81.00
Hyderabad	21.4	51.9	10	9	+ 1	1.00
Cochin	24.5	49.4	5	7	- 2	4.00
Panji	31.7	30.7	2	2	0	0.00
Etawah	21.0	66.1	11	18	- 7	49.00
Saharanpur	17.2	68.0	17.5	21	+ 3.5	12.25
Howrah	10.1	56.5	25	13	+ 12	144.00
Roorkee	16.6	78.7	20	24	- 4	16.00
Noida	29.3	33.2	3	3	0	0.00

\*Assumed data

$$\Sigma(R_x - R_y)^2 = 1199.50$$

categories. Ten percent of the population are rural members and Democrats, 25 percent are rural members and Republicans, 5 percent are rural and belong to other sub-categories. The joint contingency of the two variables can be shown in a table (Table 9.9).

**Table 9.9. Relationship between political preference and place of residence**

<i>Class</i>	<i>Democrats</i>	<i>Republicans</i>	<i>Others</i>	<i>Total</i>
Urban	30	20	10	60
Rural	10	25	5	40
Total	40	45	15	100

Under these conditions when residence is known, urban members should be predicted to be Democrats, and rural members should be predicted to be Republicans. This additional information about place of residence makes it possible to reduce the total error of prediction from 55 percent errors-when only the distribution of political preference was known-to 45 percent when the additional information was taken into account. (The 45 percent error is made up of predicting Democrats for the 20 percent Republicans and Democrats for the 10 percent others in the urban group, and predicting Republicans for the 10 percent Democrats and Republicans for the 5 percent others in the rural group.)

The coefficient of relative predictability is defined as the ratio of difference between the original error and the error after a second attribute is taken into account, divided by the original error. If  $E_1$  is the original error and  $E_2$  is the error taking into account the associated variable, the coefficient  $C$  can be expressed as :

$$\text{R. P.} = \frac{E_1 - E_2}{E_1}$$

Referring to Table 9.9, the original error,  $E_1$  is the sum of nonmodal categories for the first attribute-political preference. That

is, the marginal categories 40 plus 15, or 55 per cent. The error under the second condition,  $E_2$ , is the sum of all the non-modal, or error, cells in the contingency Table 9.9; that is, 20 plus 10, plus 10 plus 5, or 45 per cent. The improvement in prediction is 10 per cent and the coefficient of relative predictability is 10 divided by 55 or 0.189.

Since directionality is not defined for attributes-they are either present or absent-the coefficient can never be negative. If the amount of error under the second condition is equal to the original error, the coefficient is zero and it can be said that the attributes are not related. Table 9.10 has been constructed according to random probability so that the distribution within the table is exactly the chance distribution.

$E_1$  for this table is 40 plus 15 or 55 per cent and  $E_2$  is 24 plus 9, plus 16 plus 6, also 55 per cent.

$$R.P. = \frac{55 - 55}{55} = 0$$

**Table 9.10. Relation between political preference and place of residence**

Class	Democrats	Republicans	Others	Total
Urban	24	27	9	60
Rural	16	18	6	40
Total	40	45	15	100

If the attributes are related perfectly so that no error in prediction is found, the coefficient is one. Table 9.11 shows perfect predictability.

$E_1$  for this table is 45 plus zero or 45 per cent,  $E_2$  is zero, and  $R.P. = 45 - 0/45 = 1$ . The coefficient of 0.189 for the two attributes shown in Table 9 indicates some correlation between these attributes and a corresponding improvement in prediction if one of the attributes can be identified for each member for the population.

**Table 9.11. Relationship between political preference and place of residence**

<i>Class</i>	<i>Democrats</i>	<i>Republicans</i>	<i>Others</i>	<i>Total</i>
Urban	45	0	0	45
Rural	0	55	0	55
Total	45	55	0	100

Other measures of association between attributes have been proposed. Most of the better known measures are based on some application of *chi square* as a test of the null hypothesis of no association. Many of them are restricted to fourfold tables and may have other underlying assumptions or restrictions to their general usefulness.

Yule has proposed a coefficient usually designated as "*Q*" which measures association in a fourfold table. *Q* is found by computing the ratio between the difference and the sum of the cross products of the diagonal cells. That is, if the cells of the fourfold table are designated as in the figure below.

A	B
C	D

$$Q = \frac{AD - BC}{AD + BC}$$

*Q* varies between minus one and plus one as *BC* is less than or greater than *AD*. The sign of *Q* is determined by the arrangement of the cells. Frequently this arrangement is arbitrary with no meaningful interpretation in the problem. If *AD* = *BC* it can be assumed that there is no association between the variables and *Q* = 0. If either *AD* or *BC* is zero, the association is complete—that is, the correlation is perfect—and *Q* = ± 1.

Other measures of contingency include  $\phi$  which is defined as:

$\phi = \sqrt{X^2 / N}$ , and is defined as

$$C = \sqrt{\frac{X^2}{N}} \quad \text{or} \quad C = \sqrt{\frac{\phi^2}{1 + \phi^2}}$$

an equivalent formula. Tschuprow attempted to refine these contingency measures which are based on *chi square* to take into account the number of categories involved in the cross classification. Tschuprow's  $T$  is defined by the formula

$$T^2 = \frac{X^2}{N \sqrt{(s-1)(t-1)}}$$

where  $s$  is the number of rows and  $t$  is the number of columns in the contingency table. All of these measures have serious limitations due to their dependence on assumptions, restrictions to the use of *chi square*, and the dependence of the results on the proper definition of units of classification of the variables. A thorough knowledge and understanding of these concepts and assumptions is essential to the proper application of measures of contingency which are based on distribution statistics such as *chi square*.

**Curvilinear correlation :** As was already indicated, the Pearsonian or product-moment coefficient of correlation is based on the assumption that the relationship between the two variables is rectilinear. When the data are non-linear (curvilinear),  $r$  does not accurately measure the amount of correlation between the variables. The constancy of the ratio of change of the two variables determines whether the correlation is rectilinear or curvilinear. If the amount of change in the two variables bears a constant ratio, the correlation is rectilinear, but if it does not, then the relationship is curvilinear. For example, the relationship between the strength and age of human beings is curvilinear since strength does not bear a constant ratio to

age throughout the entire span of life. Facts of this kind can be brought out clearly by plotting the data on a scatter diagram.

**Partial and multiple correlation :** Since the social scientist is frequently concerned with a large number for factors and their interrelations, simple correlation techniques may be found to be definitely inadequate for certain problems. Simple correlation does not measure separately the relationship between two variables in such a way that the effects of other related variables are eliminated, nor does simple correlation reliably determine the relationship between any one variable and the combined effect of several related variables. The two problems of correlation as thus stated can be more satisfactorily dealt with by the methods of partial and multiple correlation.

Partial correlation analysis is a statistical technique for measuring the degree of relationship between two variables when the effects of certain other specified variables with which they are related are eliminated. Partial correlation has certain characteristics of the experimental method in that it is possible to study the relation between two factors when several other factors are held constant. In the social sciences it is extraordinarily difficult to conduct controlled experiments in the same manner that the physicist or chemist does in the laboratory. However, partial correlation offers to the social scientist one of the most satisfactory substitutes for controlled experimentation.

Multiple correlation analysis is a statistical technique for measuring the correlation between one variable (dependent variable) and the combined effect of a number of other variables (independent variables).

## MEASURES OF RELIABILITY AND SIGNIFICANCE

A "true" measure of a distribution, such as the arithmetic mean, would be based on all the cases in the universe. The universe value of a summarizing measure of this kind is known as a *parameter*. If the true mean weight or parameter of the male students at the Gurukul Kangari University were to be computed, every male student registered would have to be included. It will be recalled that the

mean weight of 148.74 pounds was based on a sample of 265 cases, or approximately 20 per cent of the universe. This mean of 148.74 pounds is only a probable value. The value of a summarizing measure based on a sample is known as a *statistic*. If another sample of 265 cases were selected, the mean weight would not be identical to the one derived from the first sample, but it would probably be very similar. Moreover, neither the mean of the first sample nor that of the second sample would probably coincide exactly with the true mean based on the entire universe. In other words, measures based on samples are generally larger or smaller than their corresponding parameters.

Since most statistical studies are based on samples, it is therefore important to know how closely the measures based on samples represent the parameters, and also how much variation one may expect if other samples are analyzed. The so-called measures of reliability and significance may be found of great assistance in clarifying these problems.

Measures of reliability are concerned only with fluctuations due to random sampling. Obviously, they have nothing to do with observational or computational errors. Moreover, whenever a measure of reliability is computed, it is understood that the sample is adequate and has been selected according to a rigorously scientific procedure.

According to large sample theory, the reliability of a measure, such as the arithmetic mean, depends upon (1) the number of cases in the sample and (2) the variability of the values in the sample. It is no more size of the sample. We would naturally have more confidence in the trustworthiness of a sample that is relatively large than in one that is very small.

The degree of variability of the cases in a simple also has an important influence on the reliability of the measures computed from the sample. If the cases in the sample show a pronounced scatter, a greater chance fluctuation in the measures would naturally be expected. The relationship between the size of the sample and the variability of the items is clearly indicated by the formula for the standard error of the mean. The standard error is a measure of reliability :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N - 1}}$$

The standard error of the mean weight for the sample of 265 Gurukul Kangri University is :

$$\sigma_{\bar{x}} = \frac{18.37}{\sqrt{265 - 1}}$$

$$= \frac{18.37}{16.25} \\ = 1.13$$

What does this figure of 1.13 signify, or more generally, what is the "meaning" of a standards error ? In order to follow the steps in the reasoning process leading to the concept of standard error, and in order to made precise probability statements, it is necessary to reason as follows :

1. If an infinite number of samples each consisting of 265 male freshmen were drawn, about 68 per cent of the samples would have means falling within the range  $\mu \pm 1.13$  pounds (where  $\mu$  is the universe mean)
2. If the mean weight of all university freshman males were in fact known, the probability would be approximately .68 (about two chances out of three that the mean weight of any random sample of 265 freshman males drawn from among all freshmen would lie within 1.13 pounds of the universe parameter or universe mean).

The same principles apply to other measures of standard error. The formulas for the standard error of the standard deviation and for the standard error of the median are :

$$\sigma_{\sigma} = \frac{\sigma}{\sqrt{2N - 1}}$$

$$\sigma_{\hat{M}} = 1.2533 \frac{\sigma}{\sqrt{N - 1}}$$

Because of the difficulty in stating verbally the precise meaning of statistical concepts it should be pointed out that the foregoing interpretation contains certain inaccuracies. For example, the standard error of the mean has been used as if it were based on a complete knowledge of the universe of college freshmen, whereas it was actually based on only one sample of 265 men. If further samples were taken, it is probable that this figure would be slightly different.

While every statistical measure has its own reliability formula, it should be emphasized that each such formula is subject to special qualifications, limitations, and conditions. For example, the statements made above are reasonably accurate only for samples of more than 30 cases drawn from a universe which is normally distributed.

One form of the standard error of the Pearsonian coefficient of correlation is :

$$\sigma_r = \frac{1}{\sqrt{N - 1}}$$

This formula should be used only for correlation problems involving 30 or more cases, and for testing whether a given correlation coefficient is significantly different from zero.

In the illustrative problem in which the coefficient of correlation for height and weight is computed,

$$\begin{aligned}\sigma_r &= \frac{1}{\sqrt{265 - 1}} \\ &= \frac{1}{16.25} \\ &= .0615 \quad \text{or} \quad .06\end{aligned}$$

The classical or traditional techniques of estimating sampling errors as sketched in the preceding paragraphs are frequently referred to as large sample theory. There is another theory of samples known as small theory. In small sample theory.

It will no longer be open to us to assume (a) that the random

sampling distribution of a parameter is approximately normal, or even single-humped, or (b) that values given by the data are sufficiently close to the universe values for us to be able to use them in gauging the precision of our estimates.

A "small sample" generally refers to one of less than 30 to 40 cases, although in actual practice it is difficult to draw a sharp distinction between a large sample and a small sample. As a rule, the techniques of small samples are applicable to large sampling without modification, but the reverse does not hold.

In working with small samples, the distributions of the variables become extremely important and tests of significance must be related much more carefully to the proper distributions. Normal approximations no longer result in reasonable tests, since variables may assume a number of different distributions under different conditions, the mathematics and assumptions underlying the more common of these distributions should be familiar to the researcher. Any variable which is distributed according to a "success or failure" determination, such as heads or tails on a coin, true or false on a test, or occurrence or non-occurrence of any behavior will form a binomial distribution. The ratio of squared difference between observed and theoretical frequencies to the theoretical frequency is distributed according to *chi square*. There are of course other distributions, one of the most significant of which is the "*t*" distribution.

## STATISTICAL INFERENCE

The most important development in modern statistics is the logic and techniques which make it possible to use small samples with confidence. These procedures are commonly referred to as statistical inference. Statistical inference is concerned with decision making. Unlike the informal, hit-or-miss practices of daily living, the rules and techniques in statistical inference are clearly specified. Providing that there is agreement concerning rules, two or more statisticians will arrive at the same decision. The rules and techniques of statistical inference are so designed that it is not possible to be certain about the correctness of any particular decision, but in the

long run the proportion of correct decisions is known in advance. There is no insurance that one will not make an incorrect decision concerning a population parameter from a statistic based on a sample.

There are several types of errors which may occur in statistical inference : *Type I error* refers to rejection of a hypothesis when it should not be rejected. In statistical inference, the proportion of Type I errors one is willing to make is usually stated in advance. If, for example, there is an extremely low probability (say .002) that a certain characteristic is inferred about the parameter from a measure (statistic) derived from a sample, it may be decided to reject the hypothesis that the sample occurred by chance, and the probability of being in error by rejecting this hypothesis is .002. If a hypothesis is never rejected, a Type I error will never be made. Such a practice, however, will result in Type II errors. *Type II errors* occur when a hypothesis is accepted, when it should be rejected. In actual practice it is not possible to determine whether the rejection of a hypothesis is correct or incorrect. Generally, statisticians are willing to make a Type I error 5 per cent of the time. Under these circumstances the probability of making a Type I error is .05. Thus, on the average 1 time in 20, or 5 times in every 100, the rejection of a particular hypothesis will be a mistake.

The probability with which a Type I error is risked is known as level of significance. In the preceding example, the level of significance is .05. Other common levels of significance are .01 and .001. A *statistical hypothesis* is a hypothesis. The *null hypothesis* is the most widely used type of statistical hypothesis. The rejection of a null hypothesis proceeds as the basis of determining whether the discrepancy between the observed value and the population parameter is so large that it is unlikely to be the result of chance. Such a decision is made within the context of Type I and Type II errors as gauged by the *level of significance*. For example, if there is a probability greater than .05 that a null hypothesis is correct we may fail to reject the null hypothesis. Frequently, it is desirable that level of significance be .01 or .001 before failing to reject the null hypothesis.

Another type of hypothesis is concerned with the testing of the difference between statistics from two samples. The problem arises when a difference in some measure (statistic) occurs between two samples. The probability of such a difference is computed from the sampling distribution of the difference between this particular measure. If, for example, the difference in the two measures would occur less than .05 of the time, the hypothesis that the two samples came from the same population would be rejected.

Another type of problem in statistical inference is to find *confidence limits* for particular measures. For example, if the mean income of a sample is Rs. 5,000, it may be desirable to know the interval. Rather, if the process of determining such limits were repeated hundreds of times, the confidence integral would contain the parameter 95 per cent of the time. The 99 per cent confidence limits implies that in one time in 100 the parameter would be between the limits, but, of course, the limits would be wider. The less willing we are to be wrong, the less precise must be the estimate of the interval containing the parameter.

Statistical inference never produces certainty. We do not know whether we are right or wrong in a decision to reject or not to reject a hypothesis, but we do know how often we shall be wrong in rejecting when we make a series of such decisions. Some people may object that they want certainty, not probability. Such a demand is reasonable, but impossible of realization. Even in the physical sciences as a consequence of an increasing awareness of the limitations of measurement, the emphasis on the relativity of scientific conclusions has supplanted the certainties of an earlier generation.

This is merely a brief, simple, nontechnical discussion of a few basic concepts, logic, procedures, and applications of statistical inference. The ability to utilize the principles of statistical inference with proficiency, of course, demands considerably more knowledge.

**DISTRIBUTION-FREE, OR NONPARAMETRIC, METHODS OF STATISTICAL INFERENCE**

The dependence of so much of statistical inference upon the normal distribution is somewhat justified since nonnormal distributions usually can be transformed into a normal distribution. Some data are already distributed according to known mathematical functions and are subject to analysis within these functions. However, a large part of the data with which social scientists must deal is not distributed according to known mathematical functions, and assumptions as to normality, uniformity, or some other regularity are at best open to severe criticism. During the past few decades there has been a marked development in techniques for estimating parameters and testing hypotheses which require no assumption about the form of the distribution of the population. These distribution-free, or nonparametric methods are largely based on the principle of "order statistics." By "order statistics" we mean statistics that depend only on the sequence or array of the elements of the variable and are not concerned with any distance between the first and second, second and third units, etc. The fact that the distribution between any two successive observations in an array is independent of the form of the population from which the sample is drawn underlies all of the principles of distribution-free methods of inference.

In distribution analysis the mean and standard deviation are the central concepts of position and dispersion. In dispersion. In distribution-free statistics the median is the measure of central tendency and dispersion is measured in terms of various ranges, such as the interruptible or the decile range. A distribution-free measure of association is the Spearman  $\rho$ , or rank difference coefficient of correlation discussed above.

Basic to distribution-free methods is the use of the binomial distribution. By using the simple techniques of array, counting, and probabilities under the binomial distribution, tests of position and range, confidence intervals, and many hypotheses are available which do not in any way depend upon assumptions regarding the distribution of the parent population. In the future the theory and derived techniques

of distribution-free statistics will gain in emphasis. The student, especially in the social sciences, should familiarize himself with the various distribution-free methods of analysis and their wide application to small samples and relatively poorly defined variables, so that these more general methods might be substituted when assumptions necessary for more precise statistics, such as the product-moment coefficient of correlation, standard deviation, etc., can not be reasonably inferred.

### **TIMES SERIES**

In many types of problems in social research it is frequently necessary to study changes over a period of time. For some of these problems relatively simple tabular and graphic techniques will be found very satisfactory. The vertical bar chart, the rectangular coordinate graph, and the semi-logarithmic chart represent the basic graphic forms for presenting time series.

Sometimes, however, it may be necessary to use more refined mathematical techniques to analyze certain types of temporal series adequately. Because of the limitation of space, only a very few of the most elementary techniques can be included in the hope that the student may acquire some appreciation of the problems involved in the analysis of data of this kind. In actual practice it will be found that the movements in time series may assume a number of forms. The length of the period in which the movement completes its course, the general characteristics or configuration of the movements, and the degree of regularity of the movements may evidence many variations. With reference to these points, movements may be described as :

1. **Secular or long-time trend** : Social phenomena often exhibit a definite or persistent tendency to increase or decrease over a considerable period of time.
2. **Periodic fluctuations** : Perhaps the most common type is illustrated by the seasonal variation of social phenomena with more or less regularly recurrent maxima and minima. There

also may be diurnal or weekly cycles characterized by a marked periodicity.

3. **Undulatory or cyclical movements :** These movements are wavelike in character but not definitely periodic. The so-called business cycle with its alternating periods of depression and prosperity is a good example of this type of movement.
4. **Irregular variations :** These irregular movements may be episodic or they may be fortuitous or accidental. The episodic changes may be caused by such factors as strikes, lockouts, conflagrations, earthquakes, or some other type of disaster or natural cataclysm. Episodic changes result in sharp and pronounced breaks in the variable and show no apparent tendency toward recurrence at stated intervals. Fortuitous or accidental movements are generally less pronounced than episodic changes and are not ordinarily due to some easily determined cause.

In order to illustrate a few of the more elementary techniques for analyzing time series let us consider the data in Table 9.12. A cursory examination of these data will reveal that there has been a definite upward trend in mortality from diabetes in the state of Delhi during the thirty-one year period, 1970 to 2000. This fact can be brought out more clearly by plotting the data on a chart. Three basic procedures can be used in determining the secular trend of a series of this kind : (1) by drawing freehand a line of best fit, (2) by the method of moving averages, and (3) by the method of mathematical curve fitting. The freehand method is a crude and uncertain procedure for determining the secular trend of a series since it is based merely on a general visual impression. Ordinarily this procedure should not be used except for preliminary work. The method of moving averages is superior to the freehand method but it also possesses definite limitations. The moving average is obtained by averaging consecutive groups in the series; each time one year is omitted and another year is added. For example, if a three-year moving average were calculated for the series in Table 9.12, the first step would be to obtain the mean rate for the first three years :  $(66.4 + 55.5 +$

$61.7)/3 = 61.2$ . The derived mean would be recorded for the year 1971, since it is the midpoint of the first three-year period. In order to obtain similar values for the succeeding years one rate is added and one is dropped. To obtain the second moving average in Table 9.12 the rate for 1970 is omitted and the rate for 1973 added :  $(55.5 + 61.7 + 56.5)/3 = 57.9$ . The mean, 57.9, is recorded opposite the year 1972, the midpoint of the second three-year period. This procedure is continued throughout the entire series. Moving averages can be based on intervals of varying length, depending on the original data. It is not always easy, however, to determine the optimum interval. Since the means must be centered, it is advantageous to use if possible an odd number of years for the interval.

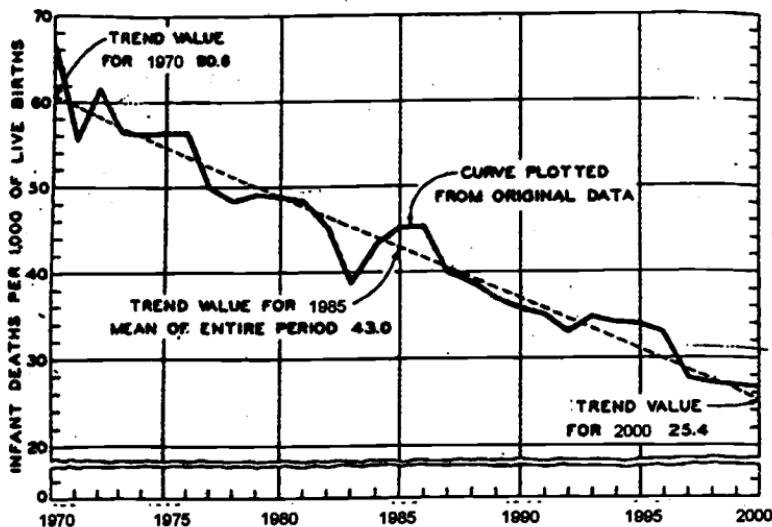
In determining a trend line mathematically, a choice has to be made of the form of line to be fitted. The line may be straight or curvilinear. Curvilinear lines may be expressed by a parabola, hyperbola, compound-interest curve, or a figure of more complicated form. For the series in Table 9.12 a straight line would seem to be the most satisfactory type. In practice it will be found that the most common type of trend is a straight line.

In Figure 9.5 a straight line of least squares has been fitted to the series of mortality rates. This type of trend line is also known as a first-degree parabola. In computing a straight line of least squares one or two common methods can be followed.

1. It will be observed from Table 9.12 that the years and the corresponding rates in the series are listed in the first and second columns.
2. The midpoint of the series is located and deviations are marked off negatively for the earlier years and positively for the latter years. In the problem in Table 9.12 the year 1975 is the mid-year or point of origin.
3. The  $y$ -values are multiplied by the corresponding deviations and the products are entered in the fourth column ( $dy$ ).
4. The deviations are squared and recorded in the fifth column ( $d^2$ ).

5. The value for the mid-year (intercept  $a$ ) is obtained by computing the mean of the values of the  $y$ -column. In the problem,  $\Sigma y/N = 1334.2/31 = 43.039$ .
6. In order to determine the slope ( $b$ ) of the line of least squares it is necessary to substitute in the following formula :  $b = \Sigma dy/\Sigma d^2$ . In the problem,  $b = -2907.9/2480 = -1.173$ . It will be seen that  $b$  may be either plus or minus, depending on whether the line of least squares shows an upward or downward movement.
7. The ordinate for any year in the series can be obtained by the formula  $y = a + bx$ . It will be recalled that  $a$  represents the mean of the  $y$ -column and is the ordinate for the mid-year of the series. In the problem,  $a = 43.039$ . In plotting a straight line of least squares on a graph it is not necessary to calculate the ordinates for each year in the series. Two ordinates, relatively widely separated, are all that are required. Let us, therefore, compute the ordinates for the first and last years in the series. For 1970,  $Y = 43.039 + (-15) (-1.173) = 60.634$  and for

Infant Mortality, Delhi 1920-1950\*



\* Assumed Data

Fig. 9.5. Fitting a straight-line trend by the method of least-squares

2000,  $Y = 43.039 + (+15)(-1.173) = 25.444$ . In Figure 9.13 the ordinates for 1970 and 2000 were plotted on the graph and then connected by a straight line.

**Table 9.12. Determination of straight-line trend by method of least squares. Data represent infant mortality rates per 1,000 live births for the Delhi State, 1970 to 2000\***

Year	Rate y	Devi- ation from Mid- point d	Devi- ation times Rate dy	Devi- ation Squared $d^2$	$b = \frac{\sum dy}{\sum d^2}$	$= \frac{-2907.9}{2480}$				
					$= -1.173$					
					1	2	3	4	5	6
1970	66.4	- 15	- 996.0	225						
1971	55.5	- 14	- 777.0	196						
1972	61.7	- 13	- 802.1	169						
1973	56.5	- 12	- 678.0	144						
1974	56.2	- 11	- 618.2	121						
1975	56.4	- 10	- 564.0	100						
1976	56.4	- 9	- 507.6	81						
1977	49.8	- 8	- 398.4	64						Trend Values for :
1978	48.1	- 7	- 336.7	49						(1) 1985 (mid-point)
1979	49.0	- 6	- 294.0	36						
1980	48.7	- 5	- 243.5	25						
1981	48.3	- 4	- 193.2	16	$a = \frac{\sum y}{N}$					
1982	45.2	- 3	- 135.6	9						
1983	38.8	- 2	- 77.6	4	$= \frac{1334.2}{31}$					
1984	43.2	- 1	- 43.2	1						
1985	45.2	0	0	0	$= 43.039$					
1986	45.4	1	45.4	1	(2) 1970					
1987	39.9	2	79.8	4	$= 43.039 + (-15)(b)$					

<i>I</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	
1988	38.7	3	116.1	9	= 43.039 + (-15) (-1.173)
1989	36.8	4	147.2	16	= 43.039 + (17.595)
1990	35.7	5	178.5	25	= 60.364
1991	35.0	6	210.0	36	
1992	33.1	7	231.7	49	(3) 2000
1993	34.9	8	279.2	64	= 43.039 + (+ 15) (b)
1994	34.1	9	306.9	81	= 43.039 + (+ 15) (-1.173)
1995	33.9	10	339.0	100	= 43.039 + (-17.595)
1996	33.0	11	363.0	121	= 25.444
1997	27.7	12	332.4	144	
1998	27.1	13	352.3	169	
1999	26.8	14	375.2	196	
2000	26.7	15	400.5	225	

\*Assumed data

## STATISTICAL ERRORS

To be really proficient in statistical work a thorough knowledge of the various techniques is, of course, indispensable, but in addition such qualities as common sense, good judgement, a healthy scepticism, objectivity, experience, and a broad understanding of the field of study are also essential. Many of the more serious errors in statistical work are of a non-mathematical character. This fact is often overlooked by the beginner who has acquired some acquaintance with statistical formulae and techniques. Statistical analysis is not a mere perfunctory mechanical process of applying formulae and operating a calculating machine. Statistical work requires good judgment, a critical attitude, and careful thought.

In order to guard against the common pitfalls of statistical work let us summarize some of the more typical sources of error : (1) inadequate and inaccurate data, (2) mechanical mistakes, and (3) unsound interpretations.

*Under the heading of inadequate and inaccurate data the following sources of error might be mentioned : (1) insufficient data, (2) unrepresentative samples, (3) data that have been deliberately falsified by informants, (4) inaccurate data resulting from poor observation and carelessness, and (5) unreliable standards and units of measurement.*

*Mechanical errors include : (1) mistakes in arithmetic and other mathematical processes, (2) application of wrong formulae, and (3) errors in copying.*

*The more common fallacies in interpretation are : (1) failing to consider all significant factors, (2) ignoring negative evidence, (3) mistaking correlation for causation, (4) comparing non-comparable data, (5) generalizing from too few cases, and (6) distorting interpretations to fit preconceptions and prejudices.*

Anyone who aspires to do statistical research should be thoroughly familiar with the four simple rules of statistical procedure which Adolph Quetelet formulated : (1) Never have preconceived ideas as to what the figures are to prove. (2) Never have reject a number that seems contrary to what you might expect merely because it departs a good deal from the apparent average. (3) Be careful to weigh and record all the possible causes of an event, and do not attribute to one what is really the result of the combination of several. (4) Never compare data which are not fully comparable.

## NOTATION AND FORMULAE

### Mean and Variance

If  $X_1, X_2, \dots, X_n$  are  $n$  values of a variable  $X$ ,

$$\text{the mean } \bar{X} = \frac{\Sigma X}{n};$$

the variance (or second moment about the mean)

$$s_x^2 = \frac{\Sigma (X - \bar{X})^2}{n} = \frac{\Sigma X^2}{n} - \bar{X}^2;$$

and the standard deviation =  $\sqrt{(\text{variance})} = s_x$ .

## Change of origin and scale

If  $X = a + kx$ , where  $a$  is an 'assumed mean' and  $x$  is the 'working variable',

$$\text{the mean } \bar{X} = a + k\bar{x}.$$

and

$$\text{the variance } s_x^2 = k^2 s_x^2 = k^2 \left( \frac{\sum x^2}{n} - \bar{x}^2 \right)$$

## Frequency distributions

If  $X_1, X_2, \dots$ , occur with frequencies  $f_1, f_2, \dots$ ,

$$\bar{X} = \frac{\sum fX}{n} \text{ and } \bar{x} = \frac{\sum fx}{n};$$

$$s_x^2 = \frac{\sum fX^2}{n} - \bar{X}^2 \text{ and } s_x^2 = \frac{\sum fx^2}{n} - \bar{x}^2$$

$$\bar{X} = a + k\bar{x} \text{ and } s_x = ks_x, \text{ as before.}$$

## Continuous distributions

If  $f(x)$  is the frequency-density for a variable  $x$ ,

the total frequency  $n = \int f(x) dx$ ;

and the mean  $\bar{x} = \int x f(x) dx$ ;

$$\text{the variance } s^2 = \frac{1}{n} \int (x - \bar{x})^2 \cdot f(x) dx$$

$$= \frac{1}{n} \int x^2 \cdot f(x) dx = \bar{x}^2$$

## Bivariate distributions

If  $(X_1, Y_1), \dots, (X_n, Y_n)$  are  $n$  pairs of values of variables  $X, Y$ , and  $(x_1, y_1), \dots, (x_n, y_n)$  are those of the 'working variables'  $x, y$ , where  $X = a + kx$  and  $Y = b + ly$ , the covariance, or product-moment about  $(\bar{X}, \bar{Y})$ ,

$$\begin{aligned}
 P_{X,Y} &= \frac{\Sigma (X - \bar{X})(Y - \bar{Y})}{n} \\
 &= kl \frac{\Sigma (x - \bar{x})(y - \bar{y})}{n} \\
 &= kl \left( \frac{\Sigma xy}{n} - \bar{x}\bar{y} \right)
 \end{aligned}$$

The regression line of  $Y$  on  $X$  is

$$Y - \bar{Y} = \frac{P_{X,Y}}{\sigma_x^2} (X - \bar{X}).$$

Correlation coefficient

$$= \frac{P_{XY}}{\sigma_x \sigma_y} = \frac{P_{xy}}{\sigma_x \sigma_y}.$$

### Probability distributions

If  $p_1, p_2, \dots$ , are the probabilities of a variable taking values  $x_1, x_2, \dots$ , where  $p_1 + p_2 + \dots = 1$ ,

the expected value  $E(x) = \mu = \sum px$ ;

and the variance  $\sigma^2 = \frac{\sum px^2}{n} - \mu^2$

If  $f(x)$  is the probability-density for a continuous variable  $x$ ,

the total probability =  $\int f(x) dx = 1$ ;

the expected value  $E(x) = \mu = \int xf(x) dx$ ;

and the variance  $\sigma^2 = \int x^2 f(x) dx - \mu^2$ ,

the integrals being taken over the possible range of values of  $x$ .

### **Binomial probability distribution**

If  $p$  is the probability of success and  $q$  that of failure in any single trial,  $p + q = 1$ , and

the probability of  $r$  successes in  $n$  trials  $P(r) = {}_n C_r q^{n-r} p^r$ ;

expected number of successes  $= np$ ;

variance  $= npq$ ; and standard deviation  $= \sqrt{(npq)}$ .

The expected proportion of successes  $= p$ , with variance  $pq/n$ , and standard deviation  $\sqrt{(pq/n)}$ .

### **Poisson probability distribution**

If  $m$  is the expected number of successes ( $= np$ , where  $n$  is large and  $p$  small),

$$P(r) = e^{-m} \cdot \frac{m^r}{r!}; \text{ variance} = m; \text{ and standard deviation} = \sqrt{m}.$$

### **Normal probability distribution**

If  $m$  is the expected value of a variable  $X$  and  $\sigma$  the standard deviation,

$$\text{the probability-density } y = \frac{1}{\sqrt{(2\pi\sigma^2)}} e^{-(x-m)^2/(2\sigma^2)}$$

or, if  $X = m + x$ ,

$$y = \frac{1}{\sqrt{(2\pi\sigma^2)}} e^{-x^2/(2\sigma^2)}.$$

### **Sampling**

If a population consists of  $N$  members, with mean  $\mu$  and variance  $\sigma^2$ , and a sample of  $n$  members,  $x_1, x_2, \dots, x_n$ , is taken, with mean  $\bar{x}$  and variance  $s^2$ ,

The mean of the sample has a probability-distribution having a Standard error of the mean  $= \sigma^2/n$

Standard error of the mean =  $\sigma/\sqrt{n}$ .

Form a large sample, estimate of mean of the population =  $\bar{x}$   
and unbiased estimate of variance =  $ns^2/(n - 1)$ .

## **STATISTICAL EXAMPLES :**

### **Problem-9.1**

A survey is to be conducted covering five institutions of Jaipur to assess the distribution of students (boys and girls separately) according to religion which has been grouped into following 4 classes (1) Hindu (2) Muslim (3) Jain (4) Others.

Prepare the form of a blank table to incorporate the above information.

**Solution :**

**Table showing the distribution of students according to religion in institutions of Jaipur**

## Alternate Table

The first table is, however, more appropriate.

### **Problem-9.2**

**Prepare a blank table to incorporate the information desired to be collected through the following questionnaire.**

- (1) Do you own a house ? Yes/No.

(2) If yes, state where it is.

  - (a) Wholly occupied by you.
  - (b) Partly rented.
  - (c) Wholly rented.

(3) If a rented house, state whether it is government quarter.  
Information is collected area-wise.

**Solution :**

**Table showing the distribution of portions according to type of houses.**

	Own House				Rented house			
Area	Wholly occupied	Partly rented	Wholly rented	Total	Govt. govt	Non govt	Total	Total

### Problem 9.3

The following report appeared about a survey of reading habits of the people.

"Out of 500 persons contacted only 224 read newspapers regularly. Others read only casually. The number of readers for different languages viz. Hindi, English, Bengali, Gujarati and others were respectively, 106, 58, 24, 10 and 26. Of the 224 persons who read regularly only 122 persons purchase their own paper, others borrow or read from other place. Language wise break up of those who purchase it is Hindi (40) English (36) Bengali (20) Gujarati, (8) others (18) Represent it in the form of a suitable table".

### Solution

Type of readers	Purchase own paper	Borrowing	Total
<b>1. Regular readers</b>			
Hindi	40	66	106
English	36	22	58
Bengali	20	4	24
Gujrati	8	2	10
Others	18	8	26
Total	122	102	224
<b>2. Casual reader</b>			
			276
	Total		500

### Problem 9.4

A survey was conducted to study the impact of smoking among the college students. 100 students addicted to smoking were selected, 50 each from Arts and Science sections. Of the 100 students addicted to smoking 40 were living with family and 60 away from it. Of the 40 students living with family 28 belonged to rich poor ones. There were 8 poor students of Arts and 18 students of Science living with family. Of the 60 students living away from their family there 20 who belonged to the poor families and out of these 12 belonged to

arts and remaining to science.

Tabulate the above information in suitable form.

### Solution

Table showing the distribution of student addicted to smoking.

<i>Faculty</i>	<i>Living with family</i>			<i>Living without family</i>			<i>Total</i>
	<i>Rich</i>	<i>Poor</i>	<i>Total</i>	<i>Rich</i>	<i>Poor</i>	<i>Total</i>	
Art	10	8	18	20	12	32	50
Science	18	4	22	20	8	28	50
Total :	28	12	40	40	20	60	100

### Problem 9.5

The following data are related to the size of family in a sample survey of 20 families. Regroup them in the form of a discrete series.

No. of persons in each of 20 families

4, 5, 2, 3, 4, 7, 2, 3, 4, 3, 5, 4, 2, 4, 3, 6, 3, 4, 7, 4.

### Solution

<i>No. of persons in each family</i>	<i>Tally Mark</i>	<i>No. of families</i>
2	III	3
3	IIII	5
4	IIIIII	7
5	II	2
6	I	1
7	II	2

### Problem 9.6

Arrange the following information in the form of continuous series with class-interval of 5

7, 10, 18, 25, 30, 25, 18, 19, 15, 27, 12, 13, 16, 0, 5, 2, 9, 8, 2, 14, 15, 19, 24, 16, 19, 3, 0, 1, 6.

**Solution**

<i>Class limits</i>	<i>Tally Marks</i>	<i>Frequency</i>
0-5		6
5-10		5
10-15		4
15-20		9
20-25	I	1
25-30		3
30-35	I	1

**Problem 9.7**

Given below are the monthly incomes of 40 students of a class. Arrange the data in the form of continuous series with class-interval of 50, (1-50, 51-100 and so on).

25, 52, 240, 250, 290, 350, 365, 190, 120, 125, 410, 180, 418, 314, 248, 260, 270, 310, 340, 337, 398, 185, 264, 346, 268, 118, 158, 175, 190, 224, 229, 345, 200, 245, 124, 230, 140, 135, 240, 285.

**Solution**

<i>Class limits</i>	<i>Tally Marks</i>	<i>Frequency</i>
1-50	I	1
51-100	I	1
101-150		6
151-200		7
201-250		8
251-300		6
301-350		7
351-400	II	2
401-450	II	2

**Problem 9.8**

The following figures relate to the heights of 10 Gorkhas selected at random. Find out Mean, Median.

Heights in centimeters 120, 126, 130, 125, 128, 136, 140, 122, 134, 132.

### Solution

<i>Serial No.</i>	<i>Height (Serially arranged)</i>
1	120
2	122
3	125
4	126
5	128
6	130
7	132
8	134
9	136
10	140
Total	1293

$$\text{Mean} = \frac{1293}{10} = 129.3$$

$$\text{Median} = \text{Size of } \left( \frac{10 \times 1}{2} \right) \text{ th item.}$$

= Size of 5.5 th item

$$= \frac{128 + 130}{2} = 129.$$

### Problem 9.9

The following are the marks gained out of 20 by 10 candidates in aptitude test. Calculate the Mean and the Median.

Marks : 8, 10, 15, 13, 12, 16, 4, 2, 6, 14.

**Solution**

<i>Serial No.</i>	<i>Marks (Serially arranged)</i>
1	2
2	4
3	6
4	8
5	10
6	12
7	13
8	14
9	15
10	16
<b>Total</b>	<b>100</b>

$$\text{Mean} = \frac{100}{10} = 10.$$

$$\text{Median} = \text{Size of } \left( \frac{10+1}{2} \right) \text{ th item}$$

= Size of 5.5 th item

$$= \frac{10+12}{2} = 11.$$

**Problem 9.10**

From the following data regarding the birth and death rates in a town for 11 years. Calculate the Mean and Median for the two separately.

Year : 1990 91 92 93 94 95 96 97 98 99 2000

Birth rate : 17.5 17.9 18.4 18.9 16.0 18.4 16.4 17.5 18.3 19.0 19.2

Death rate: 12.5 11.6 11.9 12.8 12.6 10.8 11.5 12.9 10.5 10.7 13.2

**Solution**

<i>S.N.</i>	<i>Birth rate (Serially arranged)</i>	<i>Death rate (Serially arranged)</i>
1	16.0	10.5
2	16.4	10.7
3	17.5	10.8
4	17.5	11.2
5	17.9	11.5
6	18.3	11.6
7	18.4	12.5
8	18.4	12.6
9	18.9	12.8
10	19.0	12.9
11	19.2	13.2
Total	197.5	130.3

**Mean -** Birth rate =  $\frac{197.5}{11} = 17.95$

Death rate =  $\frac{130.3}{11} = 11.82$

**Median -** Birth rate = Size of  $\left(\frac{11+1}{2}\right)$  6th item  
= 18.3

Death rate = Size of 6th item = 11.6.

**Problem 9.11**

The following table gives the data regarding the age of the mothers at the time of birth of the first and the second child. Calculate the Mean and Mode.

<i>Age in years</i>	<i>No. of mothers Birth of first child</i>	<i>No. of mothers Birth second child.</i>
15	12	-
16	14	-
17	20	3
18	35	8
19	40	12
20	30	20
21	21	46
22	14	38
23	8	22
24	5	18
25	1	12
26	-	11
27	-	10
Total	200	200

**Solution****Calculation of Mean**

<i>Size (m)</i>	<i>First child</i>			<i>Second child</i>		
	<i>Frequency ass. mean (f)</i>	<i>Dev. from (dx)</i>	<i>Frequency ass. mean (f) × (dx)</i>	<i>Frequency ass. mean (f)</i>	<i>Dev. from (dx)21</i>	<i>(fdx)</i>
15	12	- 4	- 48	-	-	-
16	14	- 3	- 42	-	-	-
17	20	- 2	- 40	3	- 4	- 12
18	35	- 1	- 3	8	- 3	- 24
19	40	0	0	12	- 2	- 24
20	30	+ 1	+ 30	20	- 1	- 20
21	21	+ 2	+ 42	46	0	0
22	14	+ 3	+ 42	38	+ 1	+ 38
23	8	+ 4	+ 32	22	+ 2	+ 44
24	5	+ 5	+ 25	18	+ 3	+ 54
25	1	+ 6	+ 6	12	+ 4	+ 48
26	-	-	-	11	+ 5	+ 55
27	-	-	-	10	+ 6	+ 60
	200	+ 12	200		+ 219	

$$\text{Mean age for the first child} = x + \frac{\sum f dx}{N}$$

$$= 19 + \frac{12}{200}$$

$$= 19.06$$

$$\text{Mean age for the 2nd child} = 21 + \frac{219}{200}$$

$$= 22.095$$

Modal age for the 1st child = 19, having highest frequency

Modal age for the second child = 21.

### Calculation of Median

Size <i>(first child)</i>	Frequency		(Second child) <i>(f)</i>	Frequency	
	<i>f</i>	Cum. Freq. <i>(c.f)</i>		<i>f</i>	Cum. Freq. <i>(c.f)</i>
15	12	12	-	-	-
16	14	26	-	-	-
17	20	46	3	3	3
18	35	81	8	11	11
19	40	121	12	23	23
20	30	151	20	43	43
21	21	172	46	89	89
22	14	186	38	127	127
23	8	194	22	149	149
24	5	199	18	167	167
25	1	200	12	179	179
26	-	-	11	190	190
27	-	-	10	200	200
	200			200	

$$\text{Median age for the first child} = \text{Size of } \left( \frac{200+1}{2} \right)$$

$$= 100.5^{\text{th}} \text{ item} = 19 \text{ years}$$

$$\text{Median age for the second child} = 21, \text{ size of } 100.5^{\text{th}} \text{ item}$$

$$= 22 \text{ years.}$$

**Problem 9.12**

A survey was conducted among 80 urban and 120 rural families to estimate the impact of family planning. The following table giving the number of children born to each couple has been drawn up. Calculate the average number of children born per pair. Use Mean, Median and Mode in all two cases.

<i>No. of children born</i>	<i>No. of couples (Urban families)</i>	<i>No. of couples (Rural families)</i>
0	4	5
1	8	12
2	12	14
3	25	17
4	10	21
5	8	13
6	5	11
7	4	10
8	3	9
9	1	8
<b>80</b>	<b>120</b>	

**Solution****Calculation of Mean**

<i>Urban families</i>				<i>Rural families</i>			
<i>(m)</i>	<i>f</i>	<i>(dx)</i>	<i>(fdx)</i>	<i>Dev. from Freq. ass. mean 3 (f) 2 (dx)</i>			
				<i>f</i>	<i>(dx)</i>	<i>(fdx)</i>	<i>Dev. from Freq. ass. mean 4 (f) × (dx)</i>
0	4	- 3	- 12	5	- 4	- 20	
1	8	- 2	- 16	12	- 3	- 36	
2	12	- 2	- 12	14	- 2	- 18	
3	25	0	0	17	- 1	- 17	
4	10	+ 1	+ 10	21	0	0	
5	8	+ 2	+ 16	13	+ 1	+ 13	
6	5	+ 3	+ 15	11	+ 2	+ 22	
7	4	+ 4	+ 16	10	+ 3	+ 30	
8	3	+ 5	+ 15	9	+ 4	+ 36	
9	1	+ 6	+ 0	8	+ 5	+ 40	
<b>80</b>			<b>+ 38</b>	<b>120</b>			<b>+ 40</b>

Mean No. of children in urban families.

$$x + \frac{\sum f dx}{N}$$

$$= 3 \times \frac{38}{80} \\ = 3.475$$

Mean number of children in rural families.

$$= x + \frac{\sum f dx}{N}$$

$$= 4 + \frac{40}{120} \\ = 4.33.$$

Modal No. of children-

Urban families = 3 (the size with highest frequency)

Rural families = 4 (the size with highest frequency)

### Calculation of Median

<i>Measurement</i>	<i>Urban Families</i>		<i>Rural Families</i>	
	<i>f</i>	<i>Cum Freq.</i>	<i>f</i>	<i>Cum. Freq.</i>
0	4	4	5	5
1	8	12	12	17
2	12	24	14	31
3	25	49	17	48
4	10	59	21	69
5	8	67	13	82
6	5	72	11	93
7	4	76	10	103
8	3	79	9	112
9	1	80	8	120

$$\text{Median for urban families} = \text{Size of } \left( \frac{80+1}{2} \right) = 40.5\text{th} \\ = 3$$

$$\text{Median for rural families} = \text{Size of } \left( \frac{120+1}{2} \right) = 60.5\text{th} \\ = 4.$$

### Problem 9.13

The following data gives the difference in the ages of 100 husbands and their wives. Calculate the mean difference and compare it with median difference.

Difference in years :	0	1	2	3	4	5	6	7
Number of couples :	7	10	15	17	18	16	11	6

### Solution

Measurement (m)	Freq. (f)	Cum. freq. (c.f)	Dev. from ass. mean 4 (dx)	(f) × (dx) (fdx)
0	7	7	- 4	- 28
1	10	17	- 3	- 30
2	15	32	- 2	- 30
3	17	49	- 1	- 17
4	18	67	0	0
5	16	83	+ 1	+ 16
6	11	94	+ 2	+ 22
7	6	100	+ 3	- 18
				- 49

$$\text{Mean} = x + \frac{\sum f dx}{N}$$

$$= 4 + \frac{-49}{100}$$

$$= 4 - .49 = 3.51$$

$$\text{Middle No.} = \frac{n+1}{2} = \frac{100+1}{2} = 50.5$$

Median = Size of 50.5th item = 4.

### Problem 9.14

The following table gives the age distribution of graduate student in a college consisting of 60 boys and 40 girls. Calculate the average age in each case (use Mean, Mode, and Median in all cases)

Age	14-15	16-17	18-19	20-21	22-23	Total
Numbers Boys	10	15	23	8	4	60
Girls	14	18	5	2	1	40

### Solution

#### Calculation of Mean

Measure- ment	Mid. value (m)	Freq. (Boys) (f <sub>1</sub> )	Freq. (Girls) (f <sub>2</sub> )	Dev. from ass. mean 18.5 (dx)	f <sub>1</sub> × dx (f <sub>1</sub> dx)	f <sub>2</sub> × dx (f <sub>2</sub> dx)
14-15	14.5	10	14	- 4	- 40	- 56
16-17	16.5	15	18	- 2	- 30	- 36
18-19	18.5	23	5	- 0	0	0
20-21	20.5	8	2	+ 2	+ 16	+ 4
22-23	22.5	4	1	+ 4	+ 16	+ 4
		60	40		- 38	- 84

$$\text{Mean age of boys} = x + \frac{\sum f_1 dx}{N}$$

$$\begin{aligned}
 &= 18.5 + \frac{-38}{60} \\
 &= 18.5 - .63 \\
 &= 17.87
 \end{aligned}$$

$$\text{Mean age of girls} = \frac{\sum f_2 dx}{N}$$

$$\begin{aligned}
 &= 18.5 + \frac{-84}{40} \\
 &= 18.5 - 2.1 \\
 &= 16.4.
 \end{aligned}$$

### Calculation of mode

$$\begin{aligned}
 \text{Boys} &= l_2 + \frac{f_1 - f_0}{f_1 - f_0 - f_2} \times i \\
 &= 17.5 + \frac{23 - 15}{46 - 15 - 8} \times 2 \\
 &= 17.5 + \frac{8 \times 2}{23} \\
 &= 18.2 \\
 \text{Girls} &= 15.5 + \frac{18 - 14}{35 - 14 - 5} \times 2 \\
 &= 15.5 + \frac{4 \times 2}{16} \\
 &= 15.5 + .5 = 16.
 \end{aligned}$$

**Note :** Since it is inclusive service  $l_1$  instead of 18 has been taken as mid-point between 17 and 18 for boys and between 15 and 16 for girls.

### Calculation of Median

<i>Measurement</i>	<i>Boys</i>		<i>Girls</i>	
	<i>Freq.</i> <i>(f)</i>	<i>Cum. Freq.</i> <i>(c.f)</i>	<i>Freq.</i> <i>(f)</i>	<i>Cum. Freq.</i> <i>(c.f)</i>
14-15	10	10	14	14
16-17	15	25	18	32
18-19	23	48	5	37
20-21	8	56	2	39
22-23	4	60	1	40

**Median age of boys**

$$\text{Middle No. } (m) = \frac{n}{2} = \frac{60}{2} = 30$$

$$\text{Median} = l_1 + \frac{i}{f} \times (m - c)$$

$$= 17.5 + \frac{2}{23} \times (30 - 25)$$

$$= 17.5 + \frac{2 \times 5}{23}$$

$$= 17.5 + .4 = 17.9$$

**Median age of girls**

$$\text{Middle No. } (m) = \frac{n}{2} = \frac{40}{2} = 20$$

$$\text{Median} = l_1 + \frac{i}{f} \times (m - c)$$

$$= 15.5 + \frac{2}{18} \times (20 - 14)$$

$$= 15.5 + \frac{2 \times 6}{18}$$

$$= 15.5 + .7 = 16.2$$

**Problem 9.15**

There are records of 42 criminals at the time of first prosecution is given below. Draw up a few frequency table in class interval of 10 and calculate Mean, Mode and Median.

Age in completed years - 40, 32, 23, 22, 15, 10, 8, 36, 25, 22, 9, 43, 12, 48, 20, 27, 57, 63, 33, 23, 39, 19, 45, 53, 19, 55, 19, 52, 58, 61, 51, 52, 35, 37, 21, 18, 21, 24, 26, 28, 30, 37.

**Solution****(1) Preparation of Frequency Table.**

<i>Class limits</i>	<i>Tally Marks</i>	<i>Frequency</i>
0-10	II	2
10-20		7
20-30		12
30-40		8
40-50		4
50-60		7
60-70	II	2

**(2) Calculation of Mean, Mode and Median.**

<i>Measure- ment</i>	<i>Mid-value (m)</i>	<i>Frequency (f)</i>	<i>Cum. freq. (c.f)</i>	<i>Dev. from 25 (dx)</i>	<i>(f) × (dx) (fdx)</i>
0-10	5	2	2	- 20	- 40
10-20	15	7	9	910	970
20-30	25	12	21	0	0
30-40	35	8	29	+ 10	+ 80
40-50	45	4	33	+ 20	+ 80
50-60	55	7	40	+ 30	+ 210
60-70	65	2	42	+ 40	+ 80
		42			+ 340

$$\begin{aligned}\text{Mean} &= x + \frac{\sum f dx}{N} \\ &= 25 + \frac{340}{42} \\ &= 25 + 8.1 = 33.1\end{aligned}$$

$$\begin{aligned}\text{Mode} &= l_1 + \frac{f_2}{f_0 + f_2} \times i \\ &= 20 + \frac{8}{7+8} \times 10\end{aligned}$$

$$= 20 + \frac{8 \times 10}{15}$$

$$= 20 + 5.3 = 25.3$$

$$\text{Middle No. (m)} = \frac{n}{2} = \frac{42}{2} = 21.$$

$$\text{Median} = l_1 + \frac{i}{f} \times (m - c)$$

$$= 20 + \frac{10}{12} \times (21 - 9)$$

$$= 20 + \frac{10 \times 12}{12}$$

$$= 20 + 10 = 30.$$

### Problem 9.16

From the following data regarding the percentage of divorce cases among the different age groups of population. Calculate the mean percentage of divorce cases.

Age	: 20-20	25-30	30-35	35-40	40-45	45-50	50-55	55-60
Population	: 250	200	180	132	84	62	50	42
Percentage of divorce cases	: 4.8	3.4	2.4	1.8	1.2	0.8	0.5	0.3

### Solution

Mean percentage in this case will be calculated by weighted averages. Percentage cases will be taken as measurement and population as weight. Age-group has only descriptive importance and will not be used for the purpose of calculation.

<i>Weight (w)</i>	<i>Measurement (m)</i>	<i>(m) × (w) mw</i>
250	4.8	1200.0
200	3.4	680.0
180	2.4	432.0
132	1.8	237.6
84	1.2	100.8
62	0.8	49.6
50	0.5	25.0
42	0.3	12.6
$\Sigma w = 1000$		$\Sigma mw = 2737.6$

$$\begin{aligned}\text{Weighted average} &= \frac{\sum mw}{\sum w} \\ &= \frac{2737.6}{1000} \\ &= 2.74\end{aligned}$$

### Problem 9.17

The following table gives the death rate in two towns. A and B for different age group. Calculate the average death rate in the two and show which town is healthier.

<i>Age group</i>	<i>Town A Population</i>	<i>Death rate per 000</i>	<i>Town B Population</i>	<i>Death rate per 000</i>
0-15	15000	24	20000	25
15-50	20000	20	52000	20
50 and above	5000	28	8000	30
	<b>40000</b>		<b>80000</b>	

### Solution

This is also a case of weight average. The death rate is measurement and population is weight.

Town A Measurement			Town B Measurement			
Weight (w)	(m)	(w) × (m) (mw)	Weight (w)	(m)	(w) × (m) (mw)	(wa) × (m) (mwa)
15000	24	360000	20000	25	500000	375000
20000	20	400000	52000	20	1040000	400000
5000	28	140000	8000	30	240000	150000
<b>40000</b>		<b>900000</b>	<b>80000</b>		<b>1780000</b>	<b>925000</b>

$$\text{Average death rate of Town A} = \frac{\sum mw}{\sum W}$$

$$= \frac{9,00,000}{40,000} = 22.5$$

$$\text{Average death rate of Town B} = \frac{1780000}{80000} = 22.2$$

From the death rates calculated above, it appears that Town B is healthier because death rate in that town is lower. But this is a wrong conclusion. The weights of two towns (Distribution of population in different age group) are not the same and the average death rates of the two towns have been calculated on different basis. In order to arrive at a correct conclusion, the weights for B towns should be same as for A town. In other words, we should calculate the standardized death rate of B town as follows :

Standard Death rate of B Town

$$= \left[ \frac{25 \times 15000 + 20 \times 20000 + 30 \times 5000}{40000} \right]$$

$$= \left[ \frac{375000 + 400000 + 150000}{40000} \right]$$

$$= \frac{925000}{40000}$$

$$= 23.125$$

By comparing the average death rate of A town with standardised death rate of B town, we find that A town is healthier than B town, because its death rate is lower.

### Problem 9.18

From the following data given, marks given by the students of a college, calculate Mean, Mode and Median.

Marks upto : 9    19    29    39    49

Number of students : 4    17    35    50    60

This is cumulative frequency table which will first of all be converted into ordinary frequency table as follows :

<i>Marks</i>	<i>Frequency</i>
0-9	4 - 0 = 4
10-19	17 - 4 = 13
20-29	35 - 17 = 18
30-39	50 - 35 = 15
40-49	60 - 50 = 10

### Calculation of Mean, Mode and Median.

<i>Measure- ment</i>	<i>Mid-value (m)</i>	<i>Frequency (f)</i>	<i>Cum freq. (c.f.)</i>	<i>Dev. from 24.5 (dx)</i>	$(f) \times (dx)$ <i>(fdx)</i>
0-9	4.5	4	- 20	- 80	
10-19	14.5	13	17	- 10	- 130
20-29	24.5	18	35	0	0
30-39	34.5	15	50	+ 10	+ 150
40-49	44.5	10	60	+ 20	+ 200
		60			+ 140

$$\text{Mean} = x + \frac{\sum f dx}{N}$$

$$= 24.5 + \frac{140}{60}$$

$$= 24.5 + 2.3 = 26.8.$$

$$\begin{aligned}
 \text{Mode} &= l_1 + \frac{f_2}{f_0 + f_2} \times i \\
 &= 19.5 + \frac{15}{13 + 15} \times 10 \\
 &= 19.5 + \frac{15 \times 10}{28} \\
 &= 19.5 + 5.4 = 24.9.
 \end{aligned}$$

$$\text{Middle No. (m)} = \frac{n}{2} = \frac{60}{2} = 30.$$

$$\begin{aligned}
 \text{Median} &= l_1 + \frac{1}{f} \times (m - c) \\
 &= 19.5 + \frac{10}{18} \times (30 - 17) \\
 &= 19.5 + \frac{10 \times 13}{18} \\
 &= 19.5 + 7.2 = 26.7.
 \end{aligned}$$

### Problem 9.19

The following table gives the ages of husbands and their wives at the time of their marriage. Calculate the mean age and the standard deviation in each case. Which of the two is subject to greater variation.

Numbers	:	1	2	3	4	5	6	7	8	9	10
Age of husbands	:	18	19	19	20	22	22	23	24	25	26
Age of wives	:	16	15	17	19	20	22	23	24	23	22

**Solution :**

**Calculation of mean and S.D. of husbands.**

<i>Age of Husbands</i>	<i>Dev. from 22 (dx)</i>	<i>Square of dx (dx<sup>2</sup>)</i>	<i>Age of Wives</i>	<i>Dev. from 22 (dx)</i>	<i>Square of dx (dx<sup>2</sup>)</i>
18	- 4	16	16	- 6	36
19	- 3	9	15	- 7	49
19	- 3	9	17	- 5	25
20	- 2	4	19	- 3	9
22	0	0	20	- 2	4
22	0	0	22	0	0
23	+ 1	1	23	+ 1	1
24	+ 2	4	24	+ 2	4
25	+ 3	9	23	+ 1	1
26	+ 4	16	22	0	0
	- 2	68		- 19	129

$$\text{Mean} = x + \frac{\sum dx}{n}$$

$$= 22 + \frac{-2}{10}$$

$$= 22 - .2$$

$$= 21.8.$$

$$\text{S.D.} = \sqrt{\frac{\sum dx^2}{N}} = \left( \frac{\sum dx}{N} \right)^2$$

$$= \sqrt{\frac{68}{10} \left( \frac{-2}{10} \right)^2}$$

$$= \sqrt{6.8 - (-.2)^2}$$

$$= \sqrt{6.8 - .04}$$

$$= \sqrt{6.76}$$

$$= 2.6$$

$$\text{Coeff. of S.D.} = \frac{\text{S.D.}}{\text{mean}}$$

$$= \frac{2.6}{21.8}$$

$$\text{Mean} = x + \frac{\sum dx}{N}$$

$$= 22 + \frac{-19}{10}$$

$$= 22 - 1.9$$

$$= 20.1$$

$$\text{S.D.} = \sqrt{\frac{\sum dx^2}{N}} = \left( \frac{\sum dx}{N} \right)^2$$

$$= \sqrt{\frac{129}{10} - \left( \frac{-19}{10} \right)^2}$$

$$= \sqrt{12.9 - (-1.9)^2}$$

$$= \sqrt{12.9 - 3.61}$$

$$= \sqrt{9.29}$$

$$= 3.03$$

$$\text{Coeff. of S.D.} = \frac{\text{S.D.}}{\text{mean}}$$

$$= \frac{3.03}{20.1} = .14.$$

Thus, there is greater variation in the ages of wives.

**Problem 9.20**

The age records of 10 High School boys and same number of girls are given below. Compute the mean age and Standard Deviation in both cases. Which case variation is greater.

Age in completed years.

Boys : 10, 11, 12, 12, 13, 14, 15, 16, 16, 18

Girls : 10, 11, 11, 10, 12, 13, 11, 12, 13, 15

**Solution**

Boys			Girls		
Age	Dev. from 13 ( $dx$ )	( $dx^2$ )	Age	Dev. from 12 ( $dx$ )	( $dx^2$ )
10	-3	9	10	-2	4
11	-2	4	11	-1	1
12	-1	1	11	-1	1
12	-1	1	10	-2	4
13	0	0	12	0	0
14	+1	1	13	+1	1
15	+2	4	11	-1	1
16	+3	9	12	0	0
16	+3	9	13	+1	1
18	+5	25	15	+3	9
		<b>63</b>			<b>22</b>
		<b>+ 7</b>			<b>- 2</b>

$$\text{Mean} = x + \frac{\sum f dx}{N}$$

$$\begin{aligned}
 &= 13 + \frac{7}{10} \\
 &= 13 + .7 \\
 &= 13.7
 \end{aligned}$$

$$\text{S.D.} = \sqrt{\frac{\sum dx^2}{N} - \left( \frac{\sum dx}{N} \right)^2}$$

$$= \sqrt{\frac{63}{10} - \left(\frac{7}{10}\right)^2}$$

$$= \sqrt{6.3 - .49}$$

$$= \sqrt{5.81}$$

$$= 2.4$$

$$\text{Coeff. of S.D.} = \frac{\text{S.D.}}{\text{Mean}}$$

$$= \frac{2.4}{13.7}$$

$$= .17$$

$$\text{Mean} = x + \frac{\sum dx}{N}$$

$$= 12 + \frac{-2}{10}$$

$$= 12 - .2$$

$$= 11.8$$

$$\text{S.D.} = \sqrt{\frac{\sum dx^2}{N} - \left(\frac{\sum dx}{N}\right)^2}$$

$$= \sqrt{\frac{22}{10} - \left(\frac{-2}{10}\right)^2}$$

$$= \sqrt{22 - 0.4}$$

$$= \sqrt{2.16}$$

$$= 1.5$$

$$\text{Coeff. of S.D.} = \frac{\text{S.D.}}{\text{Mean}}$$

$$= \frac{15}{118} \\ = .13$$

The variation in the ages of boy's is greater.

### Problem 9.21

The following are the marks given by the students of a Boy's school and girl's school. A sample of 10 students being selected in each case. In which case the standard of achievement is higher and which of the two groups is more homogeneous as regards academic achievement.

Boy's School : 10, 18, 25, 27, 14, 23, 35, 8, 19, 24

Girl's School : 16, 17, 20, 34, 23, 18, 9, 25, 15, 18

### Solution

Boy's School			Girl's School		
Measurement	Dev. from 20 (dx)	Dev. from 20 (dx <sup>2</sup> )	Measurement	Dev. from 18 (dx)	Dev. from 18 (dx <sup>2</sup> )
8	- 12	144	9	- 9	81
10	- 10	100	15	- 3	9
14	- 6	36	16	- 2	4
18	- 2	4	17	- 1	1
19	- 1	1	18	- 0	0
23	+ 3	9	18	0	0
24	+ 4	16	20	+ 2	4
25	+ 5	25	23	+ 5	25
27	+ 7	49	25	+ 7	49
35	+ 15	225	34	+ 16	256
	+ 3	609		+ 15	429

$$\text{Mean} = x + \frac{\sum f dx}{N}$$

$$= 20 + \frac{3}{10}$$

$$= 20 + .3$$

$$= 20.3$$

$$\text{S.D.} = \sqrt{\frac{\sum dx^2}{N} \left( \frac{\sum dx}{N} \right)^2}$$

$$= \sqrt{\frac{609}{10} - \left( \frac{3}{10} \right)^2}$$

$$= \sqrt{60.9 - 0.9}$$

$$= \sqrt{60.81}$$

$$= 7.8$$

$$\text{Coeff. of S.D.} = \frac{\text{S.D.}}{\text{Mean}}$$

$$= \frac{7.8}{20.3}$$

$$= .38$$

$$\text{Mean} = x + \frac{\sum f dx}{N}$$

$$= 18 + \frac{15}{10}$$

$$= 18 + 1.5$$

$$= 19.5$$

$$\text{S.D.} = \sqrt{\frac{\sum dx^2}{N} - \left( \frac{\sum dx}{N} \right)^2}$$

$$= \sqrt{\frac{429}{10} - \left( \frac{15}{10} \right)^2}$$

$$= \sqrt{42.9 - 2.25}$$

$$= \sqrt{40.65}$$

$$= 6.4$$

$$\text{Coeff. of S.D.} = \frac{\text{S.D.}}{\text{Mean}}$$

$$= \frac{6.4}{19.5}$$

$$= .328$$

The standard of achievement is higher in case of Boys. Both the groups are homogeneous, but homogeneity is slightly greater in case of girls.

### Problem 9.22

The following table gives the heights in inches of two groups of people. Which of them has greater uniformity.

Height in inches :	60	61	62	63	64	65
No. of persons Group A	5	10	14	11	7	3
Group B	7	8	8	10	9	8

### Solution

Measure- ment from 62 (dx)	(f)	Group A			Group B		
		(f) × (dx) (fdx)	(fdx) × (dx) (fdx <sup>2</sup> )	(f)	(f) × (dx) (fdx)	(fdx) × (dx) (fdx <sup>2</sup> )	
60	- 2	5	- 10	20	7	- 14	28
61	- 1	10	- 10	10	8	- 8	8
62	0	14	- 0	0	8	0	0
63	+ 1	11	+ 11	11	10	+ 10	10
64	+ 2	7	+ 14	28	9	+ 18	36
65	+ 3	3	+ 9	27	8	+ 24	72
		<b>50</b>	<b>+ 14</b>	<b>96</b>	<b>50</b>	<b>+ 30</b>	<b>154</b>

### Group A

$$\text{Mean} = x + \frac{\sum f dx}{n}$$

$$= 62 + \frac{14}{50}$$

$$= 62.28$$

$$\text{S.D.} = \sqrt{\frac{\sum f dx^2}{N} - \left( \frac{\sum f dx}{N} \right)^2}$$

$$= \sqrt{\frac{96}{50} - \left( \frac{14}{50} \right)^2}$$

$$= \sqrt{1.92 - 0.8}$$

$$= \sqrt{1.84}$$

$$= 1.3$$

$$\text{Coeff. of S.D.} = \frac{\text{S.D.}}{\text{Mean}}$$

$$= \frac{1.3}{62.18}$$

$$= .021$$

### **Group B**

$$\text{Mean} = x + \frac{\sum f dx}{N}$$

$$= 62 + \frac{30}{50} = 62.6$$

$$\text{S.D.} = \sqrt{\frac{\sum f dx^2}{N} - \left( \frac{\sum f dx}{N} \right)^2}$$

$$= \sqrt{\frac{154}{50} - \left( \frac{30}{50} \right)^2}$$

$$= \sqrt{3.08 - .36}$$

$$= \sqrt{2.72}$$

$$= 1.7$$

$$\text{Coeff. of S.D.} = \frac{\text{S.D.}}{\text{Mean}}$$

$$= \frac{1.7}{62.6}$$

$$= .027$$

There is greater uniformity in Group A.

### Problem 9.23

The following table gives the number of magazines contributed by two groups of people. Which of the two do you regard to be more homogeneous as regards reading habits. Which of the two groups has greater taste for reading.

No. of magazines contributed	0	1	2	3	4	5	Total	
No. of persons	Group A	12	14	8	3	2	1	40
	Group B	8	7	10	6	5	4	40

### Solution

Measur- ement	Group A				Group B			
	Dev. from $x = 2$		$(fdx) \times (dx)$		$(f) \times (dx)$		$(f) \times (dx)$	
	$(dx)$	$(f)$	$(fdx)$	$(fdx^2)$	$(f)$	$(fdx)$	$(fdx)$	$(fdx^2)$
0	- 2	12	- 24	48	8	- 16	32	
1	- 1	14	- 14	14	7	- 7	7	
2	0	8	0	0	10	0	0	
3	+ 1	3	+ 3	3	6	+ 6	6	
4	+ 2	2	+ 4	8	5	+ 10	20	
5	+ 3	1	+ 3	9	2	+ 12	36	
6	+ 3	1	+ 3	9	2	+ 13	36	
7	+ 3	1	+ 3	9	2	+ 13	36	
8	+ 3	1	+ 3	9	2	+ 10	30	
9	+ 1	2	+ 2	4	0	+ 0	0	
10	0	1	0	0	10	0	0	
11	- 1	1	- 1	1	10	- 10	10	
12	- 2	1	- 2	4	10	- 20	20	
13	- 3	1	- 3	9	10	- 30	30	
14	- 3	1	- 3	9	10	- 30	30	
15	- 3	1	- 3	9	10	- 30	30	
16	- 3	1	- 3	9	10	- 30	30	
17	- 3	1	- 3	9	10	- 30	30	
18	- 3	1	- 3	9	10	- 30	30	
19	- 3	1	- 3	9	10	- 30	30	
20	- 3	1	- 3	9	10	- 30	30	
21	- 3	1	- 3	9	10	- 30	30	
22	- 3	1	- 3	9	10	- 30	30	
23	- 3	1	- 3	9	10	- 30	30	
24	- 3	1	- 3	9	10	- 30	30	
25	- 3	1	- 3	9	10	- 30	30	
26	- 3	1	- 3	9	10	- 30	30	
27	- 3	1	- 3	9	10	- 30	30	
28	- 3	1	- 3	9	10	- 30	30	
29	- 3	1	- 3	9	10	- 30	30	
30	- 3	1	- 3	9	10	- 30	30	
31	- 3	1	- 3	9	10	- 30	30	
32	- 3	1	- 3	9	10	- 30	30	
33	- 3	1	- 3	9	10	- 30	30	
34	- 3	1	- 3	9	10	- 30	30	
35	- 3	1	- 3	9	10	- 30	30	
36	- 3	1	- 3	9	10	- 30	30	
37	- 3	1	- 3	9	10	- 30	30	
38	- 3	1	- 3	9	10	- 30	30	
39	- 3	1	- 3	9	10	- 30	30	
40	- 3	1	- 3	9	10	- 30	30	

**Group A**

$$\begin{aligned}\text{Mean} &= x + \frac{\sum f dx}{N} \\ &= 2 + \frac{-28}{40} \\ &= 2 - .7 = 1.3\end{aligned}$$

$$\begin{aligned}\text{S.D.} &= \sqrt{\frac{\sum f dx^2}{N} - \left(\frac{\sum f dx}{N}\right)^2} \\ &= \sqrt{\frac{82}{40} - \left(\frac{28}{40}\right)^2} \\ &= \sqrt{2.05 - .49} \\ &= \sqrt{1.56} \\ &= 1.2\end{aligned}$$

$$\begin{aligned}\text{Coeff. of S.D.} &= \frac{\text{S.D.}}{\text{Mean}} \\ &= \frac{1.2}{1.3} \\ &= .92\end{aligned}$$

**Group B**

$$\begin{aligned}\text{Mean} &= x + \frac{\sum f dx}{N} \\ &= 2 + \frac{5}{40} \\ &= 2.125\end{aligned}$$

$$\begin{aligned}\text{S.D.} &= \sqrt{\frac{\sum f dx^2}{N} - \left(\frac{\sum f dx}{N}\right)^2} \\ &= \sqrt{\frac{101}{40} - \left(\frac{5}{40}\right)^2}\end{aligned}$$

$$= \sqrt{2.525 - .015}$$

$$= \sqrt{2.51} = 1.6$$

$$\text{Coeff. of S.D.} = \frac{\text{S.D.}}{\text{Mean}} = \frac{1.6}{2.215} \\ = .75$$

Group B has greater taste for reading as the average number of magazines contributed by it is greater. The same group is more homogeneous since the degree of dispersion is less.

### Problem 9.24

Which of the following two groups is more homogeneous as regards income. Which group on the whole is economically better off.

### Solution

Income 0-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700  
No. of person

Group A	10	12	14	15	17	16	16
Group B	5	7	10	35	23	10	5

Measure- ment	Med. value	Dev. from $x = 350$ (dx)	Group A		Group B			
			(f)	$f \times dx$ ( $fdx$ )	$fdx \times dx$ ( $fdx^2$ )	$f \times dx$ ( $fdx$ )	$fdx \times f x$ ( $fdx^2$ )	
0-100	50	-300	10	-3000	900000	5	-1500	450000
100-200	150	-200	12	-2400	480000	7	-14000	280000
200-300	250	-100	14	-1400	140000	10	-1000	100000
300-400	350	0	15	0	0	35	0	0
400-500	450	+100	17	+1700	170000	28	+2800	280000
500-600	550	+200	16	+3200	640000	10	+2000	400000
600-700	650	+300	16	+4800	1440000	5	+1500	450000
			100 + 2900	3770000	100 + 2400		1960000	

**Group A**

$$\text{Mean} = x + \frac{\sum f dx}{N}$$

$$= 350 + \frac{2900}{100}$$

$$= 350 + 29 = 379$$

$$\text{S.D.} = \sqrt{\frac{\sum f dx^2}{N} - \left(\frac{\sum f dx}{N}\right)^2}$$

$$= \sqrt{\frac{37,70,000}{100} - \left(\frac{2900}{100}\right)^2}$$

$$= \sqrt{37,700 - 841}$$

$$= \sqrt{36,859}$$

$$= 191.9$$

$$\text{Coeff. of S.D.} = \frac{\text{S.D.}}{\text{Mean}}$$

$$= \frac{191.9}{379}$$

$$= .51$$

**Group B**

$$\text{Mean} = x + \frac{\sum f dx}{N}$$

$$= 350 + \frac{2400}{100}$$

$$= 350 + 24 = 374$$

$$\begin{aligned}
 S.D. &= \sqrt{\frac{\sum f dx^2}{N} - \left(\frac{\sum f dx}{N}\right)^2} \\
 &= \sqrt{\frac{19,60,000}{100} - \left(\frac{2400}{100}\right)^2} \\
 &= \sqrt{19,600 - 576} \\
 &= \sqrt{19,024} \\
 &= 137.9
 \end{aligned}$$

$$\begin{aligned}
 \text{Coeff. of S.D.} &= \frac{S.D.}{\text{Mean}} \\
 &= \frac{137.9}{374} = .37
 \end{aligned}$$

Group A is economically better off since average income in case of A is higher. Group B is more homogeneous since dispersion is less.

### Problem 9.25

Calculate the mean and the standard deviation from the following data.

Class limits	: 0-9	10-19	20-29	30-39	40-49	50-59
Frequency	: 9	12	16	29	11	4

### Solution

Measure- ment	Mid value	Frequency (f)	Dev. from $x = 34.5$ (dx)	$f \times dx$ (fdx)	$fdx \times dx$ ( $fdx^2$ )
0-9	4.5	9	- 30	- 270	8100
10-19	14.5	12	- 20	- 240	4800
20-29	24.5	16	- 10	- 160	1600
30-39	34.5	28	0	0	0
40-49	44.5	11	+ 10	+ 110	1100
50-59	54.5	4	+ 20	+ 80	1600
		80		- 480	17200

$$\begin{aligned}\text{Mean} &= x + \frac{\sum f dx}{n} = 34.5 + \frac{-480}{80} \\ &= 34.5 - 6 \\ &= 28.5\end{aligned}$$

$$\begin{aligned}\text{S.D.} &= \sqrt{\frac{\sum f dx^2}{N} - \left(\frac{\sum f dx}{N}\right)^2} \\ &= \sqrt{\frac{17200}{80} - \left(\frac{480}{80}\right)^2} \\ &= \sqrt{215 - 36} \\ &= \sqrt{179} = 13.4\end{aligned}$$

## **10**

# **COMPUTERS - ITS APPLICATION IN RESEARCH**

Computer is an electronic machine which carries out certain tasks with tremendous speed and accuracy. The word "Computer" is derived from the verb 'compute' which means 'to calculate'. Basically a computer is a device made up of electronic circuits and is capable of doing following types of work.

- (1) Carrying out arithmetic and scientific calculations at a tremendous speed;
- (2) storing and retrieving large volumes of information within seconds;
- (3) performing mathematical calculations accurately; and
- (4) performing logical operations (i.e. comparisons).

## **TWO ASPECTS OF COMPUTER HARDWARE AND SOFTWARE**

Computer has two aspect's : the Hardware and the Software. Broadly speaking the physical computing equipment is known as Hardware. Different Computer programs are used to cause the hardware to function in a desired way, These are known as software. Thus both software and hardware are actively associated ill giving out miraculous results. Software is compared to a driver and Hardware

to a motor-car. Now, only when the driver operates the engine, the car moves towards the desired destination.

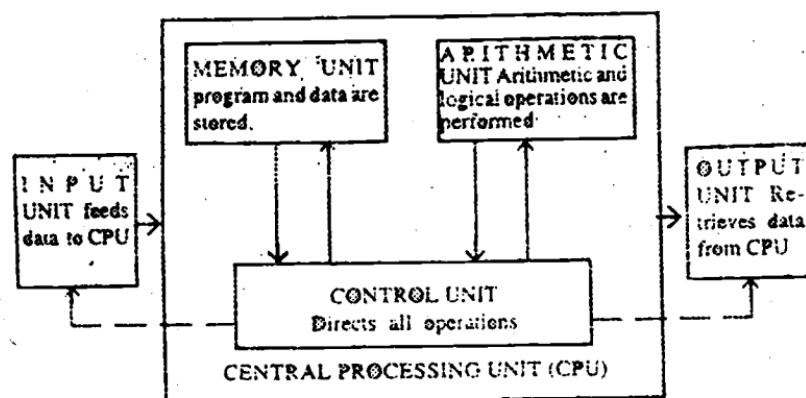
As the motor car cannot move of its own So also the computer cannot function without software or computer program. A Computer program is nothing but a set of instructions that direct a computer to perform a task.

## THE HARDWARE

In order to get an idea as to how the computer works, let us go in some more details about the Hardware of the Computer, It consists of the following :

- 1 Input Devices
- 2 Memory
- 3 Arithmetic and Logical Unit (ALU)
- 4 Output Devices
- 5 Control Unit

Input device receives the instructions in the form understandable to humans and converts it into machine usable form. The instructions after being converted into machine usable form are temporarily stored in memory.



**Fig. 10.1. Block Diagram of Computer**

The Arithmetic and Logical Unit (ALU) performs all calculations and comparisons.

Output devices take final results from memory in machine coded form and convert it into a form that can be readily used by humans.

Control Unit performs very sensitive job. It makes input unit to receive instructions in the form understandable to humans and then to convert instruction into machine usable form and then to put them in memory. The Arithmetic and Logical Unit (ALU) obtains the needed operands and carries out the operation on the operands through Central Unit. Similarly, the Output Unit obtains finished information from memory through the Control Unit. The Central Unit is therefore, aptly called the Central nervous system of the Computer.

Inter connection between the different parts of the Hardware can be illustrated in the following manner:

The Control Unit and Arithmetic and Logical Units (ALU) are combinedly called as Central Processing Unit (CPU) of the computer.

A typical computer instruction should contain the operation (for example, addition, subtraction, division, multiplication and comparison), and the operands on which the operation is to be carried out.

### **Classification of Computers**

Computers are broadly classified into three categories :

1. Digital Computer or Binary Computer
2. Analogue Computer
3. Hybrid,

### **Digital or -Binary Computers**

Digital Computers handle information as strings of binary numbers i.e, Zero (0) and Ones (1). If the operation and operand are given in binary form then the instruction is known as a machine

- instruction, Digital computer operates essentially by counting. Most computers are digital hence computers are generally called "digital computers". The digital computers are based on binary number System.

"An Arithmetic concept which uses two levels, instead of ten but operates on the same logic is called the binary system. The binary system uses two symbols "0" and "1", known as "bits", to form numbers. The base of this number system is 2. The system is called binary because it allows only two symbols for the formation of numbers. Binary numbers can be constructed just like decimal numbers except that the base is 2 instead of 10."

The two binary digits represent two states 'On' or 'Off', 'Conducting' or 'Non-conducting'. The 'On' state is represented by '1' and the 'off state by '0'. These two binary digits are called "bits". A string of right 'bits' is termed as "byte" and a group of bits constitute a word. A chip is called 8-bit, 16-bit, 32-bit and so on depending upon the number of bits contained in its standard word. All numbers are represented by the combinations of these two digits. Computers use binary system because the electrical devices can understand only "on" (1) or 'off (0).

## **Machine Language**

Different types of machine instructions and the rules governing them constitute what is known as machine language.

It is very tedious and time-consuming for the programmer to prepare programs using machine language instructions. There, an improved language known as Assembly language was developed later on. It requires , symbols to be used for operations. (For example, addition, subtraction, division, multiplication and comparison) or decimal values to be used for operands. The hardware cannot directly act upon an assembly language instruction. There has to be another program which can translate the Assembly language instructions to machine language instructions for direct execution by the computer. Such a translating program is known as an Assembler.

## High Level Languages

Even writing Assembly language program is difficult and one has to be familiar with the internal constitution and working of the hardware and the basic machine instruction set. To ease the programmer's burden, high level languages have been developed which resulted in the introduction of English-like and easy-to-use mathematical statements in the programs. For each high level language, there corresponds a translating program which translates high level statements to machine instructions. These translating programs are known as compilers.

Some of the high level languages are :

<b>FORTRAN</b>	(Formula Translation) mainly used for scientific calculations.
<b>COBOL</b>	(Common Business Oriented Language) used exclusively for business applications.
<b>BASIC</b>	(Beginner's All Purpose Symbolic Instruction Code) an interactive language use in time-sharing.
<b>PASCAL</b>	(Named after the person who developed it).
<b>ALGOL</b>	(Algorithmic Language )
<b>PL / I</b>	Programming Language-
<b>ADA</b>	Named after first Computer programmer.

**FORTRAN** - " Fortran, among other things, uses several basic statements like DO, GO TO, READ, WRITE, PUNCH, CALL, CONTINUE and IF. These instructions mean what they say; they tell the machine to do this, do that, go to this instruction, relied that instruction, and write the outcome. The power and flexibility of this seemingly simple language cannot be exaggerated. There is almost no numerical or logical operation that cannot be accomplished with it.

Fortran is a common and highly ingenious and useful communication medium. It is an intermediary language that enables the researcher, (as well as machine expert to communicate with the machine. "The researcher writes his program in Fortran, and a computer program called a compiler translates the Fortran into an equivalent machine language program. If Fortran errors are made (that is actual language and not logical errors), the compiler terminates the translation and prints out a so called diagnostic, which informs the programmer of his errors and where they are in the Fortran program."

### **Main characteristics of Computers**

**(1) Speed.** Computers can perform calculations with an unimaginable speed. "They work very fast. They are faster than almost anything else man works with. Their operations approximate the speed of light". "Research projects that would not have been attempted because of the sheer bulk of necessary calculations to analyse the data of the projects are now readily approachable with the computer and computer auxiliary equipment."<sup>4</sup>

**(2) Diligence.** "Being a machine, a computer does not suffer from the human traits of tireness and lack of concentration. If two million calculations have to be performed, it will perform the two millionth with exactly the same accuracy and speed as the first.<sup>5</sup>

**(3) Storage.** The storage of the computer is being increased continuously. As it is not possible to retain every thing in the internal memory, all unimportant information data can be stored in auxiliary storage devices. This may be brought into the main internal memory of the computer, and when required, for the processing.

**(4) Accuracy.** The computers calculate with high accuracy and dependability. It can detect errors with error detecting techniques. The errors can be human and not due to the computer.

**(5) Automation.** Computers can perform the assigned work automatically on receiving instructions.

**(6) Flexibility.** Computers can perform a job in several ways. Almost identical results can be achieved with different sets of

instructions to the machine. This flexibility of programming leads to reliability.

**(7) Ductility.** The computer machine is a faithful and objectant interpreter. It performs what a programmer tells it to do. It requires proper handling and correct and exiled input to give correct and exult output. If it is not used diligently it may create havoc.

### **Computer Generations**

Computers are not only looking with tremendous speed but also growing and developing *very fast*. A brief description of the growth and development of the computers will go to prove this. As we know, computers are of recent origin with less than fifty years to their credit. Yet during these years they have changed files and facts. This goes to prove their need and utility.

The advances in computer technology are usually described in terms of "generations." So far we have witnessed four generations of computers. The first generation computers started in 1945 and in 1990's we are expecting Fifth generation computers. Main characteristics of computer generations are given below:

**Ist Generation Computers** were those produced between 1945-60. (such as IBM-650, IBM-701.) 1st generation computers contained 1800 small bottle-sized valves which constituted its Central Processing Unit (CPU). This machine was based on vacuum tubes. It did not have any storing facility for programs and instructions had to be fed into it by a readjustment of switches and wires.

**Second Generation Computers** were those produced between 1960-65 (such as IBM-1401, Honeywell -40). These were developed after the invention of the transistor in 1947. The transistor replaced the valve use in first generation computers and made them much smaller and reliable. Thus transistors were used in the second generation Computers.

**Third Generation Computers** were produced between 1965- 70 with the invention of Integrated Circuit (IC) in 1959. The example of such computers are IBM system -360, 370, The CPU

and the main store of these computers was made of IC chips.

**Fourth Generation Computers** were produced between 1972-1990 with the invention of micro processors in 1972. The use of microprocessors as CPU in a computer has made real the dream of "Computer for the masses." Since then Microcomputers, Personal computers, Portable computers and the like have thronged the electronic market.

**Fifth Generation Computers** are in the offing. It is expected that in 1990's these will be available for use. These may use new switch (such as High Electron Mobility Transistor) instead of the present one. We are approaching fast towards use of super conducting computers. It is claimed that these computers will be 50 times more faster than the present day super fast machines. With the use of Visual Display Unit (VDU) - a T.V. like screen and a typewriter like keyboard a big strike has already been made. Magnetic tapes and discs are replaced by devices such as bubble memories and optical video discs.

### **Computers and Researchers**

From the fore-going it is clearly manifest that the computer is a very sophisticated machine. It is developing fast and its use in various fields is becoming inevitable. A Research worker cannot remain averse to it. He cannot carry on his research work perfectly and accurately without making use of Computer. He has to take help of the computer in his research. It is, therefore, necessary that he should know his computer and be able to take out work from it. He cannot depend upon the programmer in his research work. The programmer is not aware of the exact requirements of the researcher. He cannot do justice to research work. It is only the research worker who knows what is the goal and what is to be achieved and how to achieve it. It is also difficult for a programmer to study the exact nature of research and its requirement. On the other hand, the researcher can acquaint himself with the working of the computer and the facilities provided by it such as storage, memory, calculations, graphs and charts etc within a short span of time.

Computer is of immense use in research- work and data-

compilation. Analysis and comparison. In fact, we are living in the computer age. It has opened several avenues for research work, hitherto considered unimaginable and impossible. The developments taking place in computer-technology will further enhance the use of computers by the research-workers in the research-work.

## STATISTICAL SOFTWARE PACKAGES

These are now extensively used in data analysis. It is now accepted that use of a package in any Statistics Course will free students from routine and tedious hand calculations and thereby give them more time to think and understand the principles of Statistics. The chief objective of this chapter is to introduce a few of the statistical packages. In particular the features of MINITAB package are presented as relevant to the statistical procedures described in this book. It is not intended to cover all the capabilities of MINITAB, which is beyond our scope. An exhaustive list of commands, however, is in Appendix I.

Computers play an important role in statistics today. The advent of *personal computers* has changed the way statistics is applied. The availability of such facility in schools and colleges has prompted a change in the manner statistics is to be taught. In this setting, an attempt is made below to acquaint *the* reader with computer use.

Computers are well suited for statistical calculations and analysis. Writing one's own programs for this purpose using computer software languages like BASIC, PASCAL or FORTRAN is now outdated since ready-to-use programs are now available for almost all types of statistical analysis.

There is little doubt that lie computers can be very useful in learning statistics. The use of software provides an unusual opportunity for a student to learn 'how to compute', supplementing his or her understanding of 'what to compute' and 'when to compute', acquired in a formal statistics course. But when using a statistical package, it is essential to consult a statistician on its suitability and on interpretation of output.

A *statistical software package* (or statistical package for

brevity) is a group of computer programs *to perform* predetermined tasks that are often used in data analysis. These tasks include different stages comprising data organisation, presentation and analysis. This involves data entry, manipulation, graphical display and computations. Today several packages are available. We introduce three of them here; MINITAB SAS and SPSS systems. The first one is chosen for extensive illustration since it is easy to-use, quite versatile and of sufficient interest to a wide class of users in business, accounting, biological and medical research, economics, sociology and education. Moreover, it can be run on a personal computer (PC), Local Area Network (LAN) system, micro computer, or large mainframe computer.

Statistical packages are written by a few organisations or individuals and are generally known by acronyms such as BMDP, SAS, SPSS or SYSTAT. However MINITAB is not an acronym in the sense that it is not a shortened form.

A statistical package can perform on a computer most of the computations taught in lie first and second statistics courses. This allows one *to* spend more time in understanding the three W's - what, when and why *to* use a technique and, more importantly, how *to* interpret the results rather than getting bogged down by tedious arithmetic. It is essential to remember that statistical softwares and a computer are only tools. It is still you, the user, who must instruct what *to do* ! A software is no substitute for understanding the various statistical methods.

The incredible speed with which data can be analysed by a package can be wonderfully useful *to* statisticians. But this great utility can be intelligently realized, if and only if we really do know (not just think we know !) what is being computed. This means that we must compute only what we want; we should never compute some thing just because it can be computed. Especially if we do not thoroughly understand what is being computed.

## INTRODUCTION TO MINITAB : PRELIMINARIES

MINITAB is a statistical package developed at Pennsylvania

State University, U.S.A. in 1972 for use in elementary statistics courses. It is an interactive package in the sense that three user engages in a dialogue with the computer. Because of this user-friendly feature of MINITAB, errors are often detected at the time of entry and may be corrected at once. The computer on which MINITAB is run is the hardware. There are different versions of MINITAB for different types of computers. Three of these are Version II (mainly for use by students), Version 5.1 (Release 82). Version 5.1 (Release 8.2)

The MINITAB software package (MINITAB for brevity) can be installed-loaded on the hard disk (on line *memory*) or read from three disk drive. We leave it to three instructor to tell you how to activate MINITAB on your computer. MINITAB is designed to be used interactively. That is you tell MINITAB what you want it to do. It does it and then waits to be told what is to be done next. Once MINITAB is accessed by the procedure appropriate to your computer it prints out a license detail and waits to be told about the job. The display on the monitor indicates three available version of MINITAB as in the following illustration.

---

**MINITAB  
Statistical Software**

**Release 8.2**

**This software is licensed for use by :**

**MINITAB Inc.  
3081 Enterprise Drive  
State College, PA 16801-3008  
Phone: 814-238-3280**

**(C) Copyright 1991, MINITAB Inc.**

---

**Worksheet size: 16174 cells**

Press ALT+ a highlighted letter to open a menu

MTB>

In this book we use version 5.1 (Release 8.2) of MINITAB.

Once MINITAB is accessed you will see the MINITAB "prompt" - MTB >. The computer prompts you with MTB > when it is ready to receive commands through the keyboard. In fact MINITAB recognizes only the first four letters in a command; for clarity, however, it is a good practice to type it entirely. Either upper case (capital) or lower case (small) English alphabets may be used.

Typed data or commands are not entered into the computer unless the RETURN or ENTER key is pressed. Thus errors can be corrected before pressing the key. If errors still persist they may be corrected by a subsequent command. Remember always to press the RETURN or ENTER key to enter a line of data or send a command to the computer. You can come out of the MINITAB system by typing STOP. For more information on MINITAB's capabilities and commands we refer you to the MINITAB Student Hand Book, (1985) (2nd ed.) by B.F.Ryan; B.L. Jolner and T.A. Ryan.

### The Minitab Worksheet

MINITAB conducts all operations in relation to what is called a worksheet. This contains numbers in verticalities called columns. The worksheet is thus a rectangular array of numbers formed by the columns of data placed side by sides. The following provides an illustration.

Each row in Table 10.1 refers to a single student. Each column represents a different variable: Columns 1 to 5 thus refer respectively to sex (0=Female, 1=Male), income group, height, weight and journey time in minutes to reach the place of study from residence. MINITAB stores the data in exactly the same way in the computer memory. There would be a row for each case (in this example, a student) and a column for each variable. The different variables in a MINITAB worksheet are referred to by their column numbers. We put a C

**Table 10.1. Student Information**

<i>Sex</i>	<i>Group</i>	<i>Height (Cms.)</i>	<i>Weight (Kgs.)</i>	<i>Journey Time(Min.)</i>
1	1	165	64	30
0	4	158	53	28
O	1	156	52	15
1	2	170	73	42
1	3	164	66	40
0	4	155	54	35
1	4	168	70	18
1	3	164	62	20
I	2	171	75	24
0	1	153	49	18
0	1	155	60	10
0	3	150	50	15
1	1	163	60	40
1	4	170	74	50
1	2	172	76	15

before the number of the column containing the variable. For instance, to analyse journey time, we use the symbol C5 since journey time is in column 5. The limit on the number of columns (variables) and rows (cases) in a MINITAB worksheet differs from one computer to another.

#### **Format Of A Minitab Command :**

When MINITAB is instructed to do a work it looks only at four things : (i) the first four letters of the command, (ii) constants (numbers), (iii) column numbers which consist of a C followed by a number, and (iv) instruction given inside single quotes. MINITAB ignores everything else in the command. This feature allows you to give commands to MINITAB very much like you would write a statement in English. For example, if data are to be read from Table 3.1 into a MINITAB worksheet. We use the following command :

**READ the data into C1 C2 C3 C4 and C5**

Some one else might as well use the command :

**READ C1 C2 C3 C4 C5**

Both these statements are correct and MINITAB interprets them identically, since it reads only the first four letters, (which are the same in both), constants, column numbers and things in single quotes, and ignores all other words such as 'the data into' and 'and' in the first form of the command. However you should be careful about using special symbols such as and, + or # in commands since they have special uses in MINITAB.

While writing general forms of MINITAB commands we use the following conventions: Write the first four letters and the arguments that MINITAB expects. When it expects a column number write C, when it expects a constant use K and when the argument can be either a column number or a constant write E (for either). Thus for the Read command used above we write :

**READ C C ... C**

### **Reading In Data : READ command**

Data are read into a MINITAB worksheet using anyone of the two basic commands. The first is the READ command, useful for reading data consisting of several variables on a number of individuals. The data may be from a file or directly keyed in. In general the READ command is of the form

**READ C C...C**

or

**READ C - C**

The following rules must be observed while using a READ command :

**Rule 1.** When consecutive columns are to be read, the command and the arguments must be separated by at least one space (or comma).

**Rule 2.** A numerical sequence of column numbers, such

as C 1, C2, C3 and C4 may be referred briefly by using a dash (-) to mean 'from...through'.

After the command is given to read some data, MINITAB expects us to supply data. It indicates this by putting DATA > at the beginning of the line. Suppose we see on the monitor

MTB > READ data into C1-C5

DATA>

This implies that MINITAB is expecting data rather than a MINITAB command. For the READ command the data are typed in case by case. The following rules must be observed while keying in data :

- Rule 1.* When we see the first line with DATA > at the beginning, we type in the values of the variables for the first individual, with a space or comma between different values.
- Rule 2.* Numbers should. Include only a sign (+ or -), the digits and a decimal point, if necessary.
- Rule 3.* Numbers must not include any other symbols such as commas, since MINITAB considers these as more than one number. For example, the number 14,263,5 12.2 is to be typed as 142635 12.2 leaving out the commas. Otherwise three separate numbers 14,263 and 5 12.2 respectively are taken.
- Rule 4.* After the variables for the first case are typed we hit the carriage return (or enter) key. MINITAB will respond with DATA > again indicating that it expects more data. We then type in the values of the variable for the second individual. This is continued until the data are exhausted.
- Rule 5.* Hit the carriage return and MINITAB will respond with DATA. As there are no more data now, we tell this to MINITAB by typing END OF DATA (or just END) and the carriage return. MINITAB

now responds with MTB > indicating that it is waiting for a command.

- Rule 6.* The columns should have the same number of rows (i.e., the same length).

### *Reading from a stored data file :*

Data can also be read directly into the required columns by importing them from a stored file. To do this we type the command

MTB > READ data [from 'File name'] into C,...,C

The file name must be enclosed in single quotes. For example:

MTB > READ 'Mark data' into C 1-C5

### **SET Command :**

Data can also be entered by using the SET command. This enters a collection of data into one of the columns of the MINITAB worksheet. If you type, for example –

MTB > SET the following data into column C 1

DATA > 3 4 6 5 2

MTB>END

five values will be placed in column C1. The following are the rules :

*Rule 1.* The data may be typed on more than one line, if necessary.

*Rule 2.* The data values must be separated by either spaces or commas.

*Rule 3.* The completion of data entry must be indicated to the MINITAB by typing END.

Note that the following commands are equivalent to the one used above

MTB > SETC1

DATA>3 4 6 5 2

DATA >END

Several sets of data can be entered into separate columns in a MINITAB worksheet by using the SET command as many times as the number of data sets. For example, to put the four values 4, 6, 3 and 2 into column 1, five values 6, 8, 10, 11, 3 into column 2 and three values 4, 3, 5 into column 3 we type the following commands

```

MTB > SET C1
DATA> 4 6 3 2
DATA >END
MTB > SETC2
DATA> 6 8 10 11 3
DATA>END
MTB > SET C3
DATA> 4 3 5
DATA>END

```

The data values entered into MINITAB appear as follows :

#### ROW

1	4	6	4
2	6	8	3
3	3	10	5
4	2	11	
5		3	

#### Continuation of a Command

MINITAB has a special or reserved symbol to continue a command that will not fit on one line. An *ampersand* ( and) at the end of the first command line signifies that the command is continued in the next line. When MINITAB sees an ampersand it takes the next line to be a continuation of the first and does not expect a new command.

#### Naming the Columns

MINITAB commands usually have arguments that correspond

to one or more columns. In order to avoid incorrect correspondence between the variables and the columns which they stand for it is advisable to give column names. MINITAB allows naming of columns. The following are the general rules for this facility.

*Rule 1.* A column name in a command must be put in single quotes.

*Rule 2.* A column name can be almost 8 characters long.

*Rule 3.* Any character (letter, number or other symbols such as \$ or # can be used in a name. There are two exceptions: (i) A single quote cannot be used in a column name since it is used to mark the beginning and end of a name. (ii) The first and the last characters of the name cannot be blank spaces.

There can be blank spaces in the middle of a name. *The* columns can be referred to by their names or column numbers. The MINITAB output will however refer to them by name.

As an illustration, the MINITAB command for naming the five columns for the data in Table 3.1 is as follows.

NAME FOR C1 is 'SEX', C2 is 'GROUP', C3 is 'HEIGHT', C4 is 'WEIGHT', C5 is 'JTIME'

### Printing Data and Output

After the data are read into a MINITAB worksheet we may wish to check it for errors. This can be done only if the data can be displayed on the monitor or a hard copy taken. This is done by the *PRINT command* which has the general form ,

PRINT C,...,C or PRINT C - C

It is also possible to print out only one or a few columns of a worksheet. For example, we might just type

PRINT C1

to view only the contents of column C I.

Normally, MINITAB commands and output appear on the screen, but are not saved anywhere. With the commands OUTFILE

and PAPER, however. One can get a copy of the MINITAB session. To send all commands and output from session window to a printer, the PAPER command is used before starting the part of the MINITAB session to be printed. The command NO PAPER stops sending commands and output to the printer.

### **Editing the Data in a Worksheet**

Inspection of the printout of a worksheet may reveal a few errors. A single number could be incorrect or a whole row or case could have been omitted. Let us now look at the commands for correcting these errors.

#### **Correcting a Single Number: LET Command**

We must first locate the row and column in which the error lies. The row number can be located from the output of the PRINT command. We then use the following form of the LET command to replace the incorrect value.

$$\text{LET C(K1) = K2}$$

where C is the column number K1 is the row number of the incorrect value and K2 is the correct value. The *LET* command is different from most of the MINITAB commands in that it must be written precisely in the format given above. Of course, C, K1 and K2 can be replaced by the appropriate column number and constants but the parentheses and equal sign must be just as above and there can be no extra text on the line. For example, in the MINITAB session illustrated below first the SET command is used to read the height data from Table 3.1 into the third column of the MINITAB worksheet and then MINITAB is asked to print out the contents of that column. Notice that the entry in the sixth row is incorrect. It is entered as 150 instead of 155. The LET command is used to effect the correction and the third column of the MINITAB worksheet is again printed out to show the correction.

MTB > SET data into C3

DATA>165 158 156 170 164 150 168 164 171 153 155 150  
163 170 172

DATA > END of data

MTB > PRINT C3

C3

165 158 156 170 164 150 168 164 171 153 155 150 163 170  
172

MTB > LET C3 (6) = 155

MTB > PRINT C3

C3

165 158 156 170 164 155 168 164 171 151 155 150 163 170  
172

MTB>

## Defining New Variables

Sometimes we wish to study a variable which is not measured but is related to measured variables. For example, the variable journey time is in column 5 of Table 3.1. Assume that all students walk down to their place of study at an uniform speed of 6 kms per hour and we wish to find the distance between the residence and the place of study for each student. This is easily done by multiplying each number in column C5 by 6/60 or 0.1. A new variable may be defined to represent the distance travelled by the student and stored in an unused column of the MINITAB worksheet. The LET command is used here as follows.

MTB > LET C6 = 0.1 \* C5

The resulting MINITAB output is :

MTB > NAME for C6 is 'DISTANCE'

MTB > LET C6 = 0.1 \* C5

MTB > PRINT C5

DISTANCE

3.0 2.8 1.54 2.4 0 3.5 1.8 2.0 2.4 1.8 1.0 1.54 0 5.0 1.5

The general format is :

LET C= arithmetic statement

where the arithmetic statement can include one or more column numbers, constants and symbols for arithmetic operations. The most commonly used symbols are :

+	addition
-	subtraction
*	multiplication
\	division
**	raising to a power

Symbols for additional operations are listed under the LET command is Appendix A.

### **Inserting Rows: INSERT Command**

The INSERT command is used to insert a row into the worksheet. This is handy, for example, when a row was skipped while typing data. The general form of the command is

**INSERT BETWEEN ROWS K AND K OF C,....,C**

Here the first K stands for the row number in which the insertion is to be made and the second K is for the immediately preceding row. The following rule must be observed while using this command: -

**Rule :** The row numbers should be consecutive.

The English statement "We want to place the new row between the 5th and 6th rows" becomes in MINITAB :

**INSERT BETWEEN ROWS 5 AND 6 OF C,.... ,C**

If the new row is to be placed after the last row, we omit the K's all together and write

**INSERT INTO C,....,C**

After an INSERT command MINITAB will expect data and it prompts DATA>

You may type in as many lines at once as needed.

When the last line to be inserted is typed in a given location and the carriage return is pressed, MINITAB will respond with

DATA >. You then type END OF DATA to tell MINITAB that you are done, just as in the case of READ and SET commands.

The INSERT command is illustrated below using partial data from Table 10.1.

**MTB > READ the data into C 1-C3**

**DATA > 1 1 16S**

**DATA > 0 4 158**

**DATA > 0 1 156**

**DATA > 14168**

**DATA > END**

**MTB > PRINT C1-C3**

ROW	C1	C2	C3
1	1	1	165
2	0	4	158
3	0	1	156
4	1	4	168

Notice that row numbers 3,4 and 5 of data are missed while inputting the data. We correct this error by using the INSERT command:

**MTB > INSERT BETWEEN ROWS 3 AND 4 OF COLUMN C1.....,C3**

**DATA> 12170**

**DATA > 13164**

**DATA > 0 4 155**

**DATA > END OF DATA**

In the display of the contents of the columns C 1 ,C2 and C3 by a PRINT command, the inserted data will now appear after row 3 and with row numbers 4,5 and 6.

### **Deleting Rows: DELETE Command**

If a whole row is incorrect we can delete it from the worksheet by using the DELETE command with a general form :

**DELETE ROWS K,...,K of C,...,C**

The K's stand for the numbers of the rows to be deleted. Thus several rows can be deleted simultaneously by listing their numbers.

The C's stand for the numbers of the columns from which the rows are to be deleted. Thus to delete a whole row we must list all the column numbers. The following is an illustration :

**MTB > READ the data into C I-C3**

**DATA > 1 1 165**

**DATA>04 158**

**DATA > 01 156**

**DATA > 14 168**

**DATA> END**

**MTB > PRINT C 1 -C3**

ROW ;	C1	C2	C3
I	I	I	165
2	0	4	158
3	0	I	156
4	I	4	168

**MTB > DELETE ROWS 2,3 OF C I-C3**

**MTB > PRINT C I-C3**

ROW.	C1	C2	C3
1	1	1	165
2	1	4	168

**MTB > STOP**

### **Adding Comments**

Often it is desirable to add a few comments to MINITAB output so that the others easily follow what has been done. For example, if you do several assignments in one MINITAB session you may wish to add a note before each assignment so that the teacher can identify the problem sets easily. Alternatively , you

may want to record comments for your own sake so that when you look at the output later on, the variable and analysis are easily identified. The number sign, #, is used for this purpose. Whenever # appears on a line, everything after it is ignored. It can appear at the beginning of a line so that the whole line is treated as a comment or it can appear after a MINITAB command on a line, so that just the last part of the line is treated as a comment. As examples, we may wish to type the following heading on an assignment:

# HOME WORK NO.1 OF ROLL. NO.100.

# DATE OF SUBMISSION JULY 1, 1995

# PROBLEM NO.10, EX.NO.5, PAGE NO.50

Or as a reminder on our output we could type

**PRINT C4 # PRINT OUT THE WEIGHT VARIABLE IN  
COLUMN 4**

The comment type statements are not executable commands.

This implies that these are not part of the program, but they throw light on what the program does.

### Other Commands

MINITAB has several commands which provide useful information during computer sessions. We now outline some of them.

#### Information Command:

This has no arguments in the command line. It causes MINITAB to print a table of the names and row lengths of each column of data in a worksheet.

For example, an Information command on data in Table 3.1 would yield the following results .

**MTB > INFORMATION**

COLUMN	NAME	COUNT
C1	SEX	15
C2	GROUP	15
C3	HEIGHT	15

C4	WEIGHT	15
C5	TIME	15

### CONSTANTS USED: NONE

Here COUNT is the column length. The CONSTANTS USED refers to 'stored constants'. The utility of this command is that it gives a compact inventory of data without printing out each column.

### HELP Command :

We often need a brief description of a particular command in order to know the correct line format, available options or what the command does. The HELP command provides all this information. Release 8.2 has about 22 categories in HELP options. To get a list of the MINITAB commands in one of the categories, we need to type HELP COMMANDS followed by the appropriate number. The printout corresponding to some of the HELP COMMANDS with the option categories are given below:

#### (a) MTB > HELP PAPER

##### PAPER

##### Subcommands : OW OH NOTERMINAL

Normally, MINITAB commands and output appear on the screen, but are not saved anywhere. With the commands OUTFILE and PAPER, however, one can get a copy of the MINITAB session.

To send all commands and output from session window to a printer, the PAPER command is used before starting the part of the MINITAB session to be printed out. The commands OW (for output width) and OH (for output height) control the size of output. NO PAPER stops sending commands and output to the printer.

#### (b) MTB > HELP READ

READ data [from 'FILE NAME'] into C.....,C

(File > Import ASCII Data)

Read data [from 'FILE NAME.] into a K by K matrix M

(Calc > MATRICES > Read)

**Subcommands :**

**FORMAT TAB ALPHA**

**NOBS NONAMES**

reads data from a Data File or from the keyboard into columns.

By default, READ inputs data in free format with either spaces or commas as delimiters. To input data in a block format with out delimiters, use the FORMAT subcommand.

To read data from the keyboard (FILE NAME is not used), enter the data after the READ command, one line per row. For example :

**READ into C2 C3**

2 4

6 3.2

-1 6

**END**

This example puts: 2.6, and -1 in C2; and 4,3.2 and 6 in C3.

To read data from a file, the FILE NAME must be enclosed in single quotes. For example :

**READ 'MY FILE' into C1 C2**

To read data into a matrix, specify the dimensions of the matrix on the command line. For example:

**READ into a 2 by 2 matrix MI**

1 2

3 4

Note: END is not required since the matrix dimensions are specified.

**Remarks :**

1. If data are entered from the keyboard, END is to be typed after the last data line
2. Data can be entered with or without a decimal point or in

"exponential notation." Thus, the following are all equivalent.

123.4      1.234E+2      1.234E2 1234E-I

3. The missing data code (\*) can be used as data. See Help overview on the missing value code
4. With free format input, data lines may contain letters or words. These letters or words are ignored (they are not read as data). The main restriction is that the line cannot start with the first four letters of a valid command name. One cannot read alpha data using free format.
5. Stored constants cannot be used in place of numbers on a DATA line.
6. Data lines may be continued onto a new line by placing the continuation symbol and at the end of the first line.
7. The command IW (input width) controls the width of data lines input with free format. For example, with IW 72 data beyond 72 spaces will be ignored. J.W does not apply to formatted input.
8. The subcommand NOBS is not permitted when reading data into a matrix. See Help Overview on Matrices.

### (c) MTB > HELP COMMANDS

To get a list of the MINITAB commands in one of the categories below, type HELP COMMAND,s followed by the appropriate number. For example, HELP COMMANDS 5 for plotting data. The following is the full list :

1. General Information
2. Input and Output of Data
3. Editing and Manipulating Data
4. Arithmetic
5. Plotting Data
6. Basic Statistics
7. Regression
8. Analysis of Variance

9. Multivariate Analysis
10. Nonparametrics
11. Tables
12. Time Series
13. Statistical Process Control
14. Exploratory Data Analysis
15. Distributions and Random Data
16. Sorting
17. Matrices
18. Miscellaneous
19. Stored Commands and Loops
20. Design of Experiments
21. QC Macros
22. How Commands are Explained in Help

**(d) MTB > HELP Commands 1**

**General Information**

**HELP** - displays information about MINITAB

**INFO** - displays information about columns, constants  
and matrices in your worksheet

**STOP** - exits a MINITAB session

**(e) MTB > HELP Commands 2**

**Input and Output of Data**

“Data screen”..full screen data entry and editing

**READ** - puts data, row wise, into the worksheet

**SET** - puts data, column wise, into the worksheet

**INSERT** - inserts rows of data into the worksheet

**END** - signals the end of inputting data

**RETRIEVE** - retrieves a worksheet saved by MINITAB

**NAME** - assigns names to columns

- |       |  |
|-------|--|
| PRINT | - displays data from the worksheet on the screen |
| WRITE | - outputs column data to a file or the screen    |
| SAVE  | - saves the current worksheet                    |

## SAVE and RETRIEVE Commands

An advantage with computer packages is their ability to store data in the memory and retrieve it at a later time. This eliminates the need to reenter data manually each time when a different analysis is to be performed on the same data.

To store a current version of the MINITAB worksheet we simply type

MTB > SAVE 'file name'

at the end of the MINITAB program but before STOP is typed to signify the end of the session.

Once the data set has been 'saved' by MINITAB it is recorded in a format that only MINITAB can retrieve and understand. It can be made to read a previously saved data set by using the RETRIEVE command as follows: MTB > RETRIEVE 'file name' where file name is the name given to the data set when it was saved. MINITAB then creates a copy of rows and columns identical to how they appeared when they were saved. Also the data will be in the same rows and columns and with the same name as earlier. It should be noted that only the current version of the data is saved. If any changes are made in the data after retrieving them, these are not incorporated in the saved material unless that data set is saved again.

Any correction to a command statement or a data value can be made before pressing the ENTER or RETURN key by simply backspacing over the error and correcting it as required. On the other hand, if a mistype command is entered, MINITAB will recognize the error and give an error message. The error is corrected by retyping the command. However, an incorrect data entered can be corrected by using LET command.

## Ending a MINITAB Session

This is done by typing the STOP command. That is :

MTB > STOP

## ILLUSTRATIVE EXAMPLES :

### Example 10.1

Given below is an illustration of a complete MINITAB session using most of the commands introduced in this chapter. The data represent student Scores in four subjects at an examination. The first column contains a four digit student ID number; the second, third, fourth and the fifth columns, respectively, their scores in Physics, Chemistry, Mathematics and Biology. Each column is given a name and then the data are typed into the computer. After the data are printed out it was found that there was an error in row 5. A Physics score was recorded as 78 instead of 88. Also a row was missing which is to be inserted between rows 5 and 6. Finally, in order to decide whether a student has passed with distinction (70% or more in the aggregate) in first class (60- 69%), in second class (50- 59%), in pass class ( 40 - 49%) or has failed (below 40%), the total score and the percentage are calculated using two LET commands. The resulting MINITAB worksheet is printed out. The session is ended by the STOP command. MTB > READ data into C 1-C5

9 ROWS READ

MTB > END of data

MTB > NAME for C1 is 'ID NO.'

MTB > NAME for C2 is 'PHYSMARK'

MTB > NAME for C3 is 'CHEMMARK'

MTB > NAME for C4 is 'MATHMARK'

MTB> NAME for C5 is 'BIOLMARK'

MTB > PRINT C 1-C5

<b>ROW</b>	<b>ID NO.</b>	<b>PHYS MARK</b>	<b>CHEM MARK</b>	<b>KMATH MARK</b>	<b>BIOL MARK</b>
1	1210	34	53	67	56
2	1211	45	65	48	58
3	1212	49	54	53	58
4	1213	56	64	59	62
5	1214	78	79	69	78
6	1215	75	69	78	59
7	1216	56	78	76	80
8	1217	76	75	86	79
9	1218	65	49	56	60

**MTB > LET C2(S) = 88**

**MTB > INSERT between ROWS 5 and 6 OF C 1-C5**

**1 ROWS READ**

**MTB > END of Data**

**MTB > LET C6 = C2+C3+C4+C5**

**MTB > LET C7 = C6\*0.25**

**MTB > NAME for C6 is 'TOTAL'**

**MTB > NAME for C7 is 'PERCENT'**

**MTB > PRINT C1-C7**

<b>Row</b>	<b>Id No.</b>	<b>Phys</b>	<b>Chem</b>	<b>Math</b>	<b>Biol</b>	<b>Total</b>	<b>Percent</b>
1	1210	34	53	67	56	210	52.50
2	1211	45	65	48	58	216	54.00
3	1212	49	54	53	58	214	53.50
4	1213	56	64	59	62	241	60.25
5	1214	88	79	69	78	314	78.50
6	1219	57	68	65	69	259	64.75
7	1215	75	69	78	59	281	70.25
8	1216	56	78	76	80	290	72.50
9	1217	76	75	86	7.9	316	79.00
10	1218	65	49	56	60	230	57.50

**MTB > STOP**

### **Example 10.2**

We illustrate MINITAB commands for construction of a frequency table and a histogram.

### **TALLY and TABLE Commands**

An ungrouped frequency table can be obtained by using the TABLE command. This produces a table with the frequency of each different value. Subcommands can be used to give a relative frequency table along with or instead of the frequency table. The following commands produce a table with frequency of each value (specified by COUNTS) and the relative frequency as percentage (specified by the subcommand TOTPERCENTS). When there is no subcommand, the TABLE command gives the frequencies.

```
MTB > TABLE the data in column C1,  

SUBC> COUNTS;  

SUBC > TOTPERCENTS.
```

### **HISTOGRAM Command :**

This gives both a frequency table and its histogram. This has the general form:

```
MTB > HISTOGRAM of C
```

where C is the column number in which the variable of interest is stored.

When this form of HISTOGRAM command is used, MINITAB decides by itself the number of classes to be used and class width. It should be noted that the middle of the interval in the display of the histogram is not the class mark. It is a number half way between the lower class limit and 'the lower limit of the next class. MINITAB subcommand can be used to obtain a histogram with a specified class width and the first mid point of the interval. The format of the command is

```
MTB > HISTOGRAM of C ;  

SUBC>START ;  

SUBC > INCREMENT
```

The subcommands for specifying the first midpoint of the interval and the class width are respectively START and INCREMENT.

**MTB > SETC1**

**MTB > END OF DATA**

**DATA > 20 2429 212524 2625 2320 282125 26  
24 25 27 23 22 20 29 28**

**DATA > 212426 2324 2626 2527 2821 2625 24  
26 24 23 25**

**DATA >END OF DATA**

**MTB > NAME FOR C1 IS 'WEIGHT'**

**MTB>PRINTC1**

### **WEIGHT**

**20242921 2524 2625 2320 2821 25 2624 2527  
2322 20292821 2426 23 2426 2625 2728 212625  
2426 2423 25**

### **MTB>TALLYC1**

<b>WEIGHT</b>	<b>COUNT</b>
20	3
21	4
22	1
23	4
24	7
25	7
26	7
27	2
28	3
29	2
<b>N=</b>	<b>40</b>

Alternatively, one can also use the TABLE command which has sever options that can go as subcommands. For example,

MTB	>	TABLE C1 ;
SUBC	>	COUNTS;
SUBS	>	TOTPERCENT.
ROWS :		WEIGHT
		COUNT %
20	3	7.50
21	4	10.00
22	1	2.50
23	4	10.00
24	7	17.50
25	7	17.50
26	7	17.50
27	2	5.00
28	3	7.50
29	2	5.00
ALL	40	100.00

The HISTOGRAM command is illustrated below:

**MTB > HISTOGRAM C 1**

**Histogram of WEIGHT N = 40**

Midpoint	Count	
20	3	***
21	4	****
22	1	*
23	4	****
24	7	*****
25	7	*****
26	7	*****
27	2	**
28	3	***
29	2	**

In the illustration above MINITAB has assumed the starting value and the class width by default. Instead, these may be specified, as done below. This option enables the user to study the shape of

the histogram for different class widths. Thus, for example. We note that the histogram is more symmetric around the centre in the next case than in the above.

**MTB > HISTOGRAM C 1 ;**

**SUBC > START 20;**

**SUBC > INCREMENT 1.5.**

**Histogram of WEIGHT N = 40**

**Midpoint Count**

20.00	3	***
21.50	5	*****
23.00	4	****
24.50	14	*****
26.00	7	*****
27.50	5	****
29.00	2	**

**MTB > STOP**

### **3.5 OTHER PACKAGES**

#### **SAS Software**

The abbreviation SAS stands for Statistical Analysis System, developed by the SAS Institute Inc., Cary, North Carolina, USA. The system can be run on mainframe systems or on personal computers.

SAS has a much greater scope than MINITAB. It is a large collection of programs, called procedures, which may be used to analyse data sets. In contrast to MINITAB, SAS is not interactive.

#### **SPSS Software**

The Statistical Package for the Social Sciences (SPSS) was developed at the University of Chicago and the National Opinion Research Centre, primarily to aid social scientists. SPSS is the registered trade mark of SPSS Inc. 444 N Michigan Avenue, Chicago, Illinois 60611. The system has been now renamed as Statistical Presentation Software System. :

SPSS is a system that accesses and analyses data. The way you communicate with SPSS is through the SPSS command language. Commands are simply English -like statements. They can be simple, like LIST, They can contain additional specifications, such as LIST CASES FROM 1 TO 100 BY 5 or commands can contain complex combinations of subcommands, such as FREQUENCIES VARIABLES = SEX JOBCAT / BAR CHART STATISTICS ALL

The command language is easy to understand and SPSS gives you a lot of help both while learning and using it. The commands are used to do three things, usually in the following order .

1. Convert data into a form that can be used by the computer and bring them into SPSS.
2. Modify data.
3. Process data.

The SPSS package has three options to enter commands into the computer:

1. *The SPSS command prompt.* With this method, the system prompts you to enter a command; you enter it, the system evaluates and executes it.
2. *The REVIEW text editor.* This method lets you type commands into a file instead of a prompt and then submit them for execution. This method gives unlimited opportunity to edit and revise the command before execution.
3. *The Menu and Help system.* REVIEW includes a Menu and Help system that allows to select commands from menus and paste them into text editor. Along with the menus are descriptions of each command to help you figure out which one to select. The advantage of using the menu is that you don't have to remember the command names or their specifications.

The SPSS software can be run like most other statistical software packages either on a personal computer or a mainframe computer. The SESS/PC V 2.0 is the licensed version for personal computers.

SPSS represents a highly flexible package whose syntax is relatively easy to learn for individual with limited statistical and computer backgrounds. Among its features, SPSS provides excellent capabilities for variable labeling and includes a wide variety of commonly used statistical procedures.

# 11

## SCIENTIFIC SOCIAL RESEARCH

The nature and content of research can barely be comprehended without an appreciation of the method we designate as Scientific Research. Simply, it is an endeavour to arrive at answers to intellectual and practical problems through the application of scientific methods to the knowable universe. The scientific methods and procedures have been evolved and developed to increase vastly the likelihood that a class of questions viz. what? how? and to a limited extent, why? concerning the experiential phenomena will be successfully answered. Einstein and Infeld state the essence of science in the words, "Science is an attempt of the human mind to find a connection between the world of ideas and world of phenomena. All the essential ideas in Science were born in a dramatic conflict between reality and our attempt at understanding the same." This implies that science cannot always come up with an answer, let alone a definitive and correct answer, to our problems. We need not look very far to realize for ourselves that science has failed to unravel some of the persistent mysteries of human existence. Thompson observes, "The vulgar belief that science has explained everything is a hopeless misunderstanding." It should be noted that science has become relatively humble with the twentieth century revolutions, especially in physics. Today nobody is sure of being able to forecast a train of worldly events on the basis of observations, here and now. It may perhaps no longer be said that science is the only

process of knowing reality. There may well be other paths leading to this goal, but scientific procedures seem more likely to lead us there than any other method devised by man. Concentrating on science to the exclusion of other ways of approaching the world may not always be justified.

Following statement of Karl Pearson represents, by and large, one deep faith of the modern age. Pearson in his *Grammar of Science*, says "There is no shortcut to the truth,...no way to gain knowledge of the universe except through the gateway of scientific method."

He further observes, "The Scientific method is one and same in all branches (of Science) and that method is the method of all logically trained minds...the unity of all sciences consists alone in its method, not its material; the man who classified facts of any kind whatever, who sees their mutual relation and describes their sequences, is applying the Scientific method and is a man of science." These facts may belong to the history of mankind, to the social statistics of our great cities, to the atmosphere of the most distant stars, to the digestive system of a worm or to the life of a microscopic bacillus. "It is not facts themselves which make science." It is rather the method by which they are dealt with, that makes a science.

Science is independent of any particular subject matter or order of facts. "It takes the knowable universe for its subject. It deals with physical as well as psychical processes; with man as much as with nature. It has to do with everything to which its methods can be applied. What makes a science is not, of course, the nature of things with which it is concerned but the method by which it deals with these things. Study of skylark is not necessarily zoological."

There is some ambiguity about the concept of unity of science. This concept may be conceived formally as referring to logical or methodological elements common to various sciences. It has to be admitted that there are different observational bases the contents of which are explained by different sets of hypothesis and that these cannot conventionally be amalgamated. Where theory is dealt with in the abstract without regard to particular empirical contents

the logical similarity between the sciences is implicitly assumed. Methodological consideration might seem to lead to diversity rather than unity, since the operations which are actually carried out and the calculations which are actually made by workers in the various sciences seem to be insurmountably different. It is easy to exaggerate differences between the experimental situations of the physicist and the social scientist. Of course, in obvious and gross ways the meter-readings of the physicist are nothing like the interview schedules of the social scientist but the obvious and gross elements here are irrelevant to the situation. In fact, the physicist's experiment is very much like an interview with his apparatus; he asks questions of it and writes down the answers just as the psychologist does, only the questions are not verbal. "The difference between the various sciences, then are not essential differences, there is a genuine logical and methodological unity underlying their apparent diversity.

It may be helpful to distinguish between (1) the structure of science defined in terms of the rules of procedure, and (2) the corpus of science at a given time, i.e., the set of propositions accepted at a given time in accordance with the rules of procedure. Bernal says, "Science at any time is the total result of all that science has been upto date.

It should now be clear that science has a reference to a method; not so much to a field of specific subject matter. Lastrucci aptly observes, "Science is an objective, logical and systematic method of analysis of phenomena devised to permit the accumulation of reliable knowledge. It is a systematized form of analysis... not any particular body of knowledge." Aristotle was referring to this very method of dealing with facts, when he said of science that it "begins when, from a great number of experiences, one general conception is formed which will embrace all similar cases". To put it somewhat differently, science attempts to account for particular events by reference to general laws together with the actual conditions under which these laws act or to account for laws by reference to principles still more general.

The greater part of the summing up which science, accomplishes is not so much in forming new general conception as in bringing new sets of facts within the grasp of an old one. Science is an intellectual construction, a working thought-model of the world and its aim is to describe and conceptualize the impersonal facts of experience in verifiable terms, as exactly as possible, as simply as possible, as completely and meaningfully as possible. In the words of Will Durant, "Science is the captured territory (in the siege of truth). As soon as a field of enquiry yields knowledge susceptible of exact formulation, it is called Science.

Science is a human activity among others and an intelligent person is quite capable of mastering its principles. There is indeed some distinctiveness in the scientist's ways of thinking. It is, however, hard to specify what this distinctiveness is ? One would say that science is more precise. But nothing could be more precise than classical philology. Some would say that science is more concerned with truth, but nobody could be more concerned with truth than a biographer. Quite a few would point out that science looks more carefully at nature, but nobody looks more carefully at nature, than the artist. Understandably, different activities induce different habits and the longer and harder the course of study needed to acquire one set of habit, the more alienated they are likely to get vis-a-vis somebody who has been at pains to acquire a different set. It is impossible and unnecessary to learn many such sets. The physicist does not need the skills involved in collecting old archaeological pieces, nor does a literary critic need the manual dexterity involved in manipulating experimental subject; but there is no reason why each should not know in a general way the part played by different people and their *modus operandi*. We may define the three aspects of the human activity known as science, leading in effect to three objects as follows; Classification which leads to description, explanation which leads to understanding and prediction which leads to control.

In a sense, all these three aspects may be reduced to one, namely, explanation. It can be easily appreciated how classification functions quite so often as explanation. Let us take the question;

why does that animal have spots? and the answer could be, 'it is a leopard and leopards always have spots'. Take any classification; it involves such an attribution to a singular event of some universal characteristics precisely, those that mark off the member of that class from those of other classes. One may not always be satisfied with an answer of this kind and to explain why leopards have spots could warrant reference to the genetic basis of pigmentation or the evolutionary environment in which the particular animal developed. Admittedly then, the principles ultimately invoked are themselves glorified classification statements. When an event has been categorized as a case of some universal law, explanation can hardly go further. Events are explained by showing how they fit into the universe but the universe remains unexplained. Prediction can be conceived of as the application of an explanatory scheme to future events, on the assumption that if the principles hold and the conditions fulfilled which are called upon to explain some events, the event will follow. Four rules may be identified as forming a compendious summary of almost everything which is of central importance for modern Science.

*Rule I:* It is to admit no more causes of natural things than such as are both true and sufficient to explain their appearances.

*Rule II:* Therefore, to the same natural effects one must as far as possible assign the same causes.

*Rule III:* The qualities of bodies, which admit neither intention nor remission of degree and which are found to belong to all bodies within the reach of our experiments, are to be extended to the universal qualities of all bodies whatsoever.

*Rule IV:* One should look upon propositions collected by general induction from phenomena as accurately or very nearly true, notwithstanding any contrary hypothesis that may be imagined, till such time as other phenomena occur by which they may either be made more accurate or liable to exceptions.

The rules do not specify scientific method as they suggest certain restrictions which must be observed if mistakes are to be avoided. Following four rules are the rules of economy or parsimony which restricts the number of explanatory categories.

Principle of induction and restricts variety in these categories

The principle of explanation which together

Restricts scientific fantasy.

**Principle of empiricism**

Scientific method referring as it does to a procedure or a mode of investigation by which scientific and systematic knowledge is acquired, is based on certain "articles of faith." Among these are: (1) reliance on empirical evidence, (2) use of relevant concepts, (3) commitment to objectivity, (4) ethical neutrality, (5) Generality, (6) predictions based on probability and (7) public methodology affording testing of conclusions through replication. We shall discuss them in the following order.

### **1. Reliability on Empirical Evidence**

The man of Science is firmly committed to the belief that "truth" can be established on the basis of evidence that our sense organs can get at (Bertrand Russell). Of course, science never expects to reach the ultimate truths. "At their trust her theories are not and never pretend to be more than diagrams to fit, not even the possible facts, but simply the known facts (Wills). The scientist believes that the sole source of knowledge is experience (i.e., data of senses) and that there are no universal and necessary truths from which valid existential inferences can be drawn. He further believes that since knowledge of existence outside oneself is reached through experience, it must always be uncertain and tentative. All this is not to say that the scientific attitude is one of uncritical empiricism. It would be sensible to describe this attitude as critical empiricism. That is, the scientist does not accept uncritically whatever the sense datum presents itself before him. To this sense datum, he applies the screws of reason so as to comprehend its true character. In other words, the man of science regards rational ideas as the guiding principles for making predictions or formulating explanations to be tested subsequently by observation, i.e., empirical evidence, now or at some point in future. Science does not accept the derivation of a proposition from a given set of rational ideas as constituting a reliable

proof of its validity or truth. The scientist is like a creative artist who fashions a block of marble into a statue. While the insights of reason would suggest the shape and form of the statue, the artist in this process of fashioning cannot afford to remain unconcerned with the grains on and the dimensions of the marble block (empirical data) except at his own peril.

It may be regarded that the development of science as a continuing dialectical process. This implies no commitment to any special version of dialecticism, it simply takes account of the fact that what is required for the advancement of science is a continuing interplay between its logical frontiers (rationalism) and its experimental frontiers (empiricism). The logical aspect is embodied in the doctrine which has generally been known as rationalism. Rationalism proceeds from the rational investigation of connections between concept without special regard to the adequacy of the concepts of experience, developing formal structures in a free and creative way. Empiricism in doctrine, proceeds from an empirical investigation of connections between events, without special regard to the significance of those events in any total scheme of things, accumulating factual information in a disciplined and receptive manner. Both these aspects are absolutely essential and scientific progress may be regarded as a dialectical process of reciprocal feedback between them.

If empirical findings are logical constructions (theories, laws) science is at a loss; logical construction would have to catch up before the new empirical findings can be put in their place. Conversely, if logical constructions go ahead of empirical investigation that may not be regarded as so serious, because there will always be a scope for something to come up in the empirical realm to fill the new branch of logical development and provide an interpretation for part of the structure that was not interpreted before. But till such a time as it does so the logical construction is bound to remain a mere exercise of intellectual ingenuity.

The rationalists followers of old interpreted science as a deductive system of propositions. For them, there stood at the head of the system, a set of self-evident propositions and from these

other propositions (theorems) were to be derived by the process of reasoning. At the other end are the avowed inductionists (empiricists) who believe that science must construct its axioms from the sense data related to particulars by ascending continually and gradually till it finally arrives at the most general axioms.

Science operates on the twin wheels of deduction and induction, both equally germane to the goals of science. Deduction involves inferring from the premises or general statement, some bits of information about the world. Deduction is a device for the discovery of the truth that lies concealed within a set of statements. In fact, there is nothing new in deduction; all information contained in the conclusion are already contained in the premises. Nevertheless, it helps us to know and understand the world about us since it opens our eyes to the information that viewed otherwise, we would not get at. But the method of deduction is definitely limited by the facts ascertained empirically. The empirical method of extending one's palm out of the window to see if it is raining, has the advantage of rendering us safe from false premises. But the advantage of the deductive method in the instant case is that one does not have to go out and get wet to arrive at the answer. It is to be noted that the deductive method is a method of getting information just as the empirical method of getting information. In a sense, established facts have more claims to be called knowledge than inferences arrived deductively. Whenever a deduction and an empirically established fact collide the deduction must yield to the power of empirical fact. The empirical test is the final test, as someone has said; many a beautiful theory has been slain by an ugly fact. The case in the point may be variously illustrated. A plane that theoretically would not fly at all did fly despite the deduction to the contrary; the theory, the basis for deduction, in consequence, had to be revised for it was in error. The conflict between deduction and empirical knowledge cannot, however, be settled so easily. Often the empirical facts are not so clear because measurements are uncertain. In such a situation, a strong deductive argument can be more persuasive.

If one of the major aims of science is explanation, the most

usual pattern of explanation in science is evidently deductive; from a universal statement or statements (laws or principles) together with some particular statement of conditions (which together comprises the explanations) is deduced a statement describing the event to be explained. The criteria for sound explanation of this sort are that the deduction should genuinely involve the universal statements and that those statements and the statement conditions should be true as nearly as this can be ascertained.

Induction, on the contrarily, moves from particulars to arrive at general propositions. It operates on faith that in the course of things for a long time is a basic and regularity evidenced surely enough for the inference that it will continue so in the future. Induction is thus a leap of faith. Many a philosopher has indicated the paradox of induction, pointing out that past experience can barely be a secure guide to learning about nature of bodies. Their secret nature and consequently all their effect and influences may change without any change in their sensible qualities. If this happens sometimes and with regard to some object, it will happen always with regards to all objects, they point out. And then there is no logical or process argument that would secure us against this supposition. It is not inconceivable that new evidence might be forthcoming sometime and this would be only way in which the theory of induction could escape the paradox. It may nevertheless be difficult to imagine what might constitute this new evidence.

In case the premise and conclusion, in the logical case, are both known, some probability relations may be established between them and this may serve as a paradigm of an inductive inference. But where the inductively arrived at prediction has not yet been observed, where the conclusion is not known, the situation is a kin to trying to guess where the rest of the triangle lies, if one is given one side. Without further information the task is impossible, the only way to get such information is to wait. In absence of any other principle, we use, of course, the relation defined by previous sequence of observations but that the new case will conform to the pattern cannot be known until it has already done so. If we must not act

except on certainty (not probabilities) we ought not to act on religion, for it is not certain; but there are many things we do on uncertainty, sea voyages, battles etc. When we are working for tomorrow we are doing so on uncertainty, but we do not act unreasonably; for tomorrow we work for an uncertainty according to the doctrine of chance. Induction has an importance to us and hence, we are more sympathetic to proposals for providing it with some logical foundation. But the truth or falsity of the principle of induction is not altered by such efforts, any more than the truth or falsity of the existence of God is. Electing one side or the other, as a result of logical calculation, is futile any way.

The best attitude to induction is to make induction of the subject of a resolve that in the absence of any better guide to future behaviour, we would use the lessons of past experience. It would be absurd to pretend that we need reassurances about the course of events in the distant future, just as to pretend that we know anything about the course of events in the distant past. Scientific observations have been made with some accuracy for perhaps 5,000 years; they have been made in quantity and variety only for about 500 years. An extrapolation on inductive grounds into the past suggests that these periods represent an almost infinitesimal fraction of the whole life of the universe. Further, all these observations have been trade within a very thin spherical shell surrounding one planet of a small star (Sun). It could be that an animal species thus restricted in time and space has, in fact, succeeded in discovering the principles according to which the universe operates, but were it not for the fact that human beings as ourselves are member of this species, we should find the *a priori* probility of this rather small. What success we can claim lies in our constructing a theoretical account of an hypothetical universe which, supposing it existed, would be like our universe in these places and at those times where he latter has been observed. We expect that in limited predictions, the fit of the theoretical universe to the real one shall still be fairly close. Saying something beyond this would be presumptions.

The extreme empiricist view of the matter is that laws are

arrived at by induction, often understood as, by simple enumeration. But here the problem of induction is due to arise because there is no satisfactory way of explaining empirically how we can come to a position "in all cases of acts or events" and not, all observed cases of acts or events. But the failure of philosophers to solve the problem of induction has not prevented scientists from discovering laws. The fact is that the process of reasoning by which these laws are arrived at are not of induction at all. In fact they start with universal propositions as hypothesis and when they have tested them, regard them as laws.

The hypothetic reasoning runs as follows :

1. C is observed.
2. But C would follow only if A were true.
3. Therefore, there is reason that A is true.

This is the reasoning by which scientists often arrive at propositions of universal kind.

Question is often asked what the method of science is; whether induction or deduction? The only answer to this is; both. Larrabee scores the point when he remarks, "If extreme rationalist (deductionist) is like a spider spinning out theories from within, the extreme empiricist (inductionist) is to be compared, ..to an ant which piles useless heaps of facts. Better than either the spider or the ant is the bee, which selectively gathers pollen and transforms it into honey..." We must remember that in actual scientific practice, induction and deduction are mingled in intricate ways. None could have put it better than Auguste Comte who said, "Induction for deduction with a view to construction..." .

## **2. USE OF RELEVANT CONCEPTS**

Concepts are logical constructions or abstractions created from sense impressions, precepts and experiences. Concepts are the symbols that science works with; they constitute the linguistic apparatus of science. The language of science evolves in order to deal with the problems of nature for which ordinary language has proved inadequate and wanting.

The world in which we live, and in which science is discovered at work, are of apparent nature. The world which science describes is a creation of the human intellect which, while it may bear some resemblance to causal nature, is not identical with it. Neither of these taken by itself is adequate to be considered in the role of the nature which is referred to in the definition of science. Science while is the explanation of nature in its own terms, is not the explanation of apparent nature simply. What is explained is of course discovered within apparent nature. Were it not for this, we could have no access to it. But in order to be explained it is rendered, even at the descriptive level, in characteristically scientific terms, and to that extent given entry into a new realm. Explanation, being a logical relationship, lies entirely within the fields of thought and language. The nature which is explained is given in perception, but rendered in conceptual and linguistic terms. The nature in whose terms the explanation is provided, on the other hand, is not given at all, but conjectured. There are, of course, events and processes to which, for one reason or another, we cannot gain access. These constitute causal nature, having a directly productive relationship with apparent nature.

The scientific procedure consists in evolving, defining and manipulating concepts or symbols with a view to contributing variously to the established corpus of systematic knowledge and/or to establish some new bit of knowledge. In his passage from concrete sense-data to the higher and higher levels of abstraction (hypothesis, theories and laws), the man of science is constantly formulating, relying on and using relevant concepts.

### **3. COMMITMENT TO OBJECTIVITY**

The subjective-objective dichotomy is very old, going back in the history of thought beyond the foundation of most of the social and behavioural sciences. In basic outline, this dichotomy suggests that there are two fundamentally opposite ways of theoretically treating man and his social organization. One is the objective way, which views man and human society as basically similar to other aspects of the physical world. But social sciences typically prove too hazardous grounds in reference to which object frame of reference

as the right one for scientific knowledge is not totally acceptable. The object frame of reference has proved immensely useful for the physical sciences and it is not surprising, given the success of the physical sciences, that many have attempted to use this frame of reference to order and explain human behaviour. Unfortunately, human behaviour often does not tend itself to the explanation used in the physical sciences. Human behaviour entails elements which may be called ideational, i.e., intentions, meanings, values and beliefs that cannot be described in terms of sensory dimensions. The scientific method with its emphasis on objectivity bristles with problems in social sciences because of their direct or indirect concern with the study of man and his social organisation. Human behaviour can be studied by other human observers alone and they are always likely to distort the facts being observed. These facts in turn can be appreciated only on the intentional frame of reference which implies a lot of subjectivity thrown in.

The nature of scientific method is such that a practitioner of science must set aside the subjective considerations; he must be prepared to suppress his hopes and his intuition. The adoption of scientific approach may sometimes be painful but must be accorded recognition.

The man of science is firmly committed to the belief that to go nearer to the goal of truth, he must "above all things... strive at self-elimination in his judgements and provide an argument which is as true for each individual mind as his own."

Objectivity according to Galtung is consisted of

- (a) intra-subjectivity;
- (b) inter-subjectivity.

The test of intra-subjectivity (or reliability) is that repeated observations of a constant phenomenon by the same observer will yield constant data while the test of inter-subjectivity consists in finding that repeated observations of a constant phenomenon by different observers yield constant data. Inter-subjectivity is only a more adequate formulation of what is generally meant by the

“objectivity” in science. What is here involved, is not only the freedom from personal or cultural biased or partiality, but even more fundamentally the requirement that the knowledge-claims of science be in principle capable of test (confirmation or disconfirmation, at least indirectly and to some degree) on the part of any person properly equipped with intelligence and the technical devices of observation or experimentation. The term inter-subjective stresses the social nature of the scientific enterprise. If there be any “truths” that are accessible only to privileged individuals, such as mystics or visionaries, that is, knowledge-claims which by their very nature cannot be independently checked by anyone else, then such “truths” are not of the kind we seek in the sciences. The criterion of inter-subjective testability thus delimits the scientific from the non-scientific activities of man.

The scientist is thus, expected to avoid at all costs what Francis Bacon termed the “false idols.” Social sciences present typical difficulties when it comes to translating into action, the pious wish of commitment to objectivity. The critics have made much of this, some even going to the extent of insisting that the social sciences in view of their dubious objectivity would not qualify as sciences in the true sense of the term.

#### 4. ETHICAL NEUTRALITY

What Faraday said of the ‘philosopher’ applies with equal force to the scientist. “He should be a man willing to listen to every suggestion but determined to judge by himself. He should not be biased by appearances; have no favourite hypothesis; be of no school, and in doctrine have no master. He should not be a respecter of persons, but of things. Truth should be his primary object. A man of science is wedded to the faith that affectivity or commitment to an ideology is likely to distort his perspective and his judgement of things may thus become partisan or value-laden. He certainly cannot afford the luxury of prejudices, that is, believing what is comforting to believe. As Schroedniger says, Science never imposes anything, science states. Science aims at nothing but making true and adequate statements about its objects.

Generally social sciences are called upon to explain aspects of human life, it is natural that these would be sensitive to any discussion about values and moral questions. The argument for value-neutrality in social sciences makes out a case in support of it as follows.

"In order to discover, "what is and to properly conceptualize what it is, it is essential for social scientist to bring no personal prejudice or bias to his study". This does not mean that they should cease to be moral men, but for the purpose of description, for one's desire to know what is, one must observe, describe and theorise dispassionately. If disinterestedness is not maintained, what one believes may get in the way of Dogma's and would interfere with thought.

The position on ethical neutrality arose curiously among those who adopted a subjective approach to social problems. It was felt that if proper understanding of social structure, processes and behaviour demanded inference from data and an interpretive appreciation of abstract human relations, value freedom was essential. For the data to be obtained only in this way, the observer would have to hold his feelings in check for the duration of his observation and conceptualization. Since all data collection was subjective in nature, if there were no attempts at ensuring ethical neutrality, the social scientific ventures would surely raise controversies of opinion. In sum, it was thought that the social scientist should describe things as they are, to the best of his ability, keeping moral values out. He needed techniques that would actually measure things he wants to measure and not fool himself by measuring something else but this kind of argument eventually led to a new attack on the proper goal of social theorising, one which blurs the distinction between explaining something and altering it.

The criticism tends to point out that explaining things as they are amounts to putting an emphasis on the forces leading to stability and status quo and to distract people away from what might be possible by way of improvement. People who argue in this vein have often impugned the motives of social science theorists, arguing

in effect, that ethical neutrality or value free interpretations, of the social states or being were given by them in a calculated effort to justify them and keep them that way. Thus, the attack on value-neutrality usually ends up by advocating some biased view point in social analysis.

The objective of social theory is simply to explain what people do and to deduce these explanations from descriptive data organised into concepts then the problem of value-neutrality does not really arise because regardless of what one's values are in regard to the subject-matter, the same results will continue to surface. If on the other hand, explaining also means understanding or having insights into situations, perhaps in some unique human terms, then, the value problems will arise. When this happens, the distinction between social theory and bias gets blurred. One becomes deliberately biased, then, at the risk of damage to the accuracy of his results. It but this risk is sometimes worth the price in terms of the quality of insights rendered possible. This is a more palatable stance for a numerically dominant section of the social scientists today.

## 5. GENERALITY

The conclusions of any important science are generalizations; i.e., statements of general applicability. Typically, a series of observations of some class of objects, say X, are made by the scientist with a view to determining whether or not the members/items of this class have some property, say, Y. The result of these observations may be a series of protocol sentences, 'This X is Y' and so on. To avoid confusion, the scientist tries to identify the Xs in some way so as to keep them distinct from one another so that the sentences read; X<sub>2</sub> is Y', 'X<sub>n</sub> is Y'. If among a large number of such observations no X is found which is not Y and also that no X-like objects are known which exhibit great variety in Y-like properties, the scientists tend in such situations to jump from the collection of singular statements about X<sub>1</sub>, X<sub>2</sub> - X<sub>n</sub> to a universal statement about the class of Xs, viz., all Xs are Y. Such a leap is a generalization and the statement resulting from it, an empirical generalization. Generalizations emerge naturally after a large enough number of particular observations.

There can be no science without a belief in the inner harmony of the world and in the fact that reality may be grasped with the abstract theoretic or general constructions. Say Einstein and Infeld, "This belief is and will always remain the fundamental motive for all scientific creation. Throughout our efforts, in every dramatic struggle between old and new views, we recognise the eternal longing for understanding, the over firm belief in the harmony of our world continually strengthened by the increasing obstacles to comprehension.

The scientist is constantly aware of his obligation to discover "under the surface level of diversity, the thread of uniformity. Around a discovered uniformity, a logical class is constructed; about the class and its observed pattern a descriptive generalisation is formulated. Scientists are alert to opportunities for combining comparable classes into a broader class and for formulating a wider and more abstract generalization to comprehend the discrete generalizations thereby embraced," thus, the scientific theories and propositions are generated. This is what Francis Bacon suggested, precisely this when he presented his new method *Novum Organum*. Bacon advocated the method of constructing axioms from senses and particulars by ascending continually and gradually till the most general axioms are finally arrived at.

It is obvious that sciences differ in respect of the levels of generalization attained. More mature a science, the greater its generalizing potential. This has been conveyed with amazing felicity by Medawar. He says, "... the factual burden of a science varies inversely with its degree of maturity. As a science advances, particular facts are comprehended within, therefore in sense annihilated by general statements of steadily increasing explanatory powers and compass. In all sciences we are being progressively relieved of the burden of singular instances - the tyranny of the particular. We need no longer record the fall of every apple."

## 6. PREDICTIONS BASED ON PROBABILITY

The principal aspects of the scientific activity are classification which leads to description, explanation which leads to understanding

and prediction which leads to control. The human attempt to anticipate and therefore control events relies on the ability of science to predict, i.e. to obtain knowledge of future events.

Prediction is just a special type of generalization; one from the past to the future. Prediction is always a leap of faith for there is not guarantee that tomorrow will be like today. It is one's judgement and depth of knowledge about the subject matter which lends support to the supposition based on what happened in the past that something will take place in the future. A prediction is reasonable to make if our assumption is sensible that the past and future belong to the same continuum, i.e. the conditions which held in the past will obtain in the future, too. "The prediction that the sun will come up tomorrow morning is implicitly a statement that tomorrow morning comes from the same universe as have all mornings in the past.

Reliable predictions can well be made even when changes in conditions are bound to occur, if one knows the important conditions that created the trend are changing in a certain way. Since the past is never a guarantee of the future and prediction is not just a mechanical extrapolation the safer basis of projections of an observed trend into the future is an understanding of the various forces that underlie the process. Prediction shares this aspect with all generalization in prediction naturally depends on the scientist's ability to trace out the sequence of propositions embodies in the general principle faster than nature traces out the sequence of causes, so that the scientist gets there first.

The man of science believes that predictions about phenomena are possible and must rest on a solid basis of the trend repeatedly observed and the probability that very same trend would manifest itself in terms of some concrete results. The attempt to anticipate events and hence, to control them relies on the ability of science to predict. Predictions cannot be derived by deduction from any "self-evident" or "ultimate" truths. The tincture of science liberates man from the load of prejudices. Without it, the world tends to appear definite, obvious, common objects rouse no questions and familiar possibilities get contemptuously rejected. It is thus clear that the

scientific expectations or predictions are grounded in the established knowledge about the order among facts. It is well to bear in mind that these expectations may not always come true. If they do not, the scientist is under an obligation to research the corpus of knowledge or theory which initially afforded the basis for predictions and suitably amend it or even reject it. It is a part of the scientific attitude that pronouncements of science do not claim to be certain but most probable on present evidence. Probability reflects a state of the mind, best characterized not negatively as his ignorance of the future but positively as his expectation with respect to it. As Feynman says, "Scientific knowledge is a body of statements of varying degrees of certainty, some most unsure, some nearly sure, none absolutely certain.

It is typical of social sciences that they have a far lower predictability compared to the natural ones. The reasons are obviously the complexity of subject matter, inadequacy at control, etc. It is often said of the social sciences that the predictions made by them are hedged with so many preconditions (such as the well known, "other things being equal, (*Ceteris Paribus*) when they generally are not that they) are denuded of any practical worth. The much heard of distinction between the "exact" and "inexact" science stems from this though the usage 'exact' science is itself tautological, since all sciences are as exact as possible. It has now been generally agreed that the social sciences though relatively inexact are "Sciences" nevertheless and that central criterion of conferring scientific status on any branch of study should rightly be its method of study rather than the nature of results it comes out with. In other words, a science would refer to the branch of study which has progressed to a point where its analysis reveals a logical structure that is, its categories of classification, definitions and rules of correspondence are as free as possible from vagueness and ambiguity. Given time, social sciences might also vastly improve their predictive powers.

## 7. PUBLIC METHODOLOGY AFFORDING TESTING OF CONCLUSIONS THROUGH REPLICATION

Science is a public institution practising a public methodology. A scientist has to make known to others how he arrived at the conclusions he did. This way alone can the scientist expose his own methods and conclusions of his research to critical scrutiny. Criticism, according to Karl Pearson, is "the very life-blood of science." It is through such criticism alone that science as an on-going historical institution continually improves upon the means and methods of inquiry - an obligation that every true scientist shares with the rest.

Furthermore, such criticisms signal at the right moment the drawing of unwarranted conclusions which in turn might bring about considerable harm considering the fact that we, as of now depend so very much on the products of science. Science is a collective, co-operative endeavour geared , to the discovery of facts and it is, as pointed out by Dewey, "a method of knowing which itself-corrective in operation, that learns from failures as from successes.

Unless the method of scientific inquiry is made public, it would not be possible fur the fellow scientists (and critics) to replicate the initial inquiry to verify if the same conclusions are reached by recourse to the method in question. Subsequent replications lend added credence and support to conclusions of inquiries if these replications arrive at the same conclusions (assuming, of course, that similar mistakes of method do not get repeated). These are the solid bedrocks on which the corpus of science rests and from which it advances into many directions. As has just been pointed out the frequentative verification of conclusions is a basic requirement of science. This requirement brings to the fore one of the most central aspects of research; for etymologically research means repetitive searches. Such repetitive searches may lend confirmation to established conclusions in the field, help propose certain modifications in them or even invalidate them. We will do well to remember that invalidation, no less than verification of propositions is an important contribution to science.

A word about replication or repeatability since the repeatability criterion cannot be smoothly applied in the realm of social sciences.

The requirement of scientific research in regard to replication may be simply laid down as follows. The researcher must describe his empirical work in such a way that other people could repeat exactly what he did. The problem is here. The more the observer interprets what he has seen the less repeatable the study. Understandably, there is much scope in social sciences for the researchers to interpret their observations before recording them for subsequent presentation. Hence the subjective or impressionistic elements may be so strong that replication in the desired sense is not possible. For example, different observers may come to different assessments of the kind of people certain tribals are, owing to different impressions gained by them in the course of their living with them.

Modern science in contradistinction to ancient science is characterised by a measure of tentativeness with which it holds its conclusions. New data may invalidate them any moment. Developed science has removed the dogmatic arrogance of those who have never travelled the region of liberating doubt. It has kept alive our sense of wonder by showing familiar things in unfamiliar aspects, and contexts. Frequentative tests or verifications are a necessary condition for this.

It was pointed out at the outset that for a proper appreciation of the nature and content of research a thorough understanding of the scientific method is called for. In the preceding pages the salient features of scientific method were discussed at some length. An understanding of the scientific method for a student of theory and practice of research was though necessary in as much as research is, as Best puts it, "the more formal, systematic, intensive process of carrying on the scientific method of analysis." Formal aspects of the scientific method will become clearer as the following pages gradually unfold the steps involved in doing research.

## MEANING OF RESEARCH

The Webster's International Dictionary proposes a very inclusive definition of research as "a careful, critical inquiry or examination in seeking facts or principles; diligent investigation in order to ascertain

something." The above definition while helpful in indicating in a very general way what people mean when they talk about research is not specific and precise enough to afford us a thorough and clear idea of what research involves. As suggested earlier, the activities that go by the name of research involve mainly a 're-search', i.e., activities undertaken to repeat a search. Thus, perhaps improving upon the definition presented above, it may be said that research refers to "a critical and exhaustive investigation or experimentation having as its aim the revision of accepted conclusions in the light of newly discovered facts."

The researcher is constantly concerned with researching the accepted conclusions of his field, i.e., the theories with differing levels of generality and degrees of confirmation (trustworthiness) existing at a given point of time. He does this researching by probing for facts of the empirical world that confirm one or several predictions generated by his accepted conclusions, his acceptance, a consequence of his assumption about the correctness of the existing theories. Thus, researching may in effect turn to the construction of new theories to take the place of those no longer able to fit the data of the empirical world. Research, stated otherwise, is a systematic attempt to push back the bounds of comprehension and seek beyond the horizons of our knowledge some "truth" or some reality, shrouded in a subtle way and consequently, to keep on extending as also consolidating these horizons without end.

D. Slesinger and M. Stephenson in the *Encyclopedia of Social Sciences* propose a very comprehensive definition of research: "the manipulation of things, concepts or symbols for the purpose of generalizing to extend, correct or verify knowledge, whether that knowledge aids in construction of theory or in the practice of an art."

#### **(a) Manipulation of Things, Concepts or Symbols**

Sciences deal with things. In a laboratory, the physicist deals with things, e.g., balls, the psychologist deals with say, white mice. In order to know how things respond to or change under specific

conditions, what the common denominators discernible in their responses under varied conditions are and what law they abide by etc., the scientist subjects them to purposeful control, e.g., starving the rats for a specific period or letting the balls roll on different degrees of incline, etc. Such purposeful handling we call manipulation, which is an aspect of experimentation. We cannot here afford to miss out on an important point that the scientist dealing with things is not dealing with things only but also with concepts or symbols, Corresponding to things or phenomena. For the things that a particular science deals with, it has and must have some terms. As pointed out earlier, the terms designating the things about which a science tries to make sense are its concepts. If the scientist were dealing with things only, he would be working and be constrained to remain at the level of the concrete and his results would at best be limited to the particular thing in a specific situation investigated by him and none else, i.e., this ball, this rat (which again does not make much sense, considering the fact that verily, our thought process depends on concepts). Thus, the scientist dealing with things is also at the same time dealing with concepts which symbolize them and their various properties. The physicist, in our example, is thus dealing with the abstract notions of spheres, circles, velocity, acceleration, angles, pluses and minuses. The psychologist dealing with rats is dealing with abstract notions of organism, hunger, deprivation and problem-solving potential, etc. All these abstract notions relate to things that are being dealt with but in themselves are not things; these are simply the mental shorthands for things. The scientist can carry them in his head, or on a piece of paper; not so the things. Since concepts or symbols are abstract forms denuded of content, it is as though the scientist working with them was working with certain classes of things at a higher level of abstraction. Concepts have thus the relevance of things but not their bothersome content and load. The astronomer cannot get the planets and stars into his laboratory nor can he get at them and yet he can deal with them, i.e., conduct controlled inquiries into their movements because he knows the value of the variables.

It should now be clear how the concepts or symbols are subjected to manipulation just as things can be, only with greater facility and effect. Research involves thus, manipulation of things and/or concepts or symbols.

### (b) For the Purpose of Generalizing

The manipulation or purposeful control of things, concepts or symbols illustrated above is undertaken with a definite purpose, the purpose being to arrive at statements of generality. That is, the net result of a controlled inquiry should be a proposition or a conclusion which tells us to expect something to happen to a class of things when a certain class of conditions is influencing them. Of course, generality is a matter of degree and the conclusions arrived at on the basis of observations under exercise of manipulation will vary in their degrees of generality. The level of generality may be high or low; the significance or worth of a research may perhaps be judged in such terms, but absence of generality cannot characterize science. "This thing worked with Mr. X today" is no wise a conclusion of research since the concern with generalization is definitionally central to research. Slesinger and Stephenson would consider an automobile mechanic or a physician as espousing the role of a researcher to the extent each attempts to generalize about all the automobiles and all patients of a given class, respectively.

### (c) To Extend, Correct or Verify Knowledge

Generalizations thus drawn have, obviously, certain effects for the established corpus of knowledge. A general proposition having been established as an outcome of research may extend the bounds of knowledge existing at a point in time. It may, if it does not fit a certain body of existing knowledge in the sense of not auguring well with the expectations that can be legitimately drawn from it, exert pressure on the scientist to bring about certain amendments or modifications in this body of knowledge in view of the discordant note struck by this new general proposition. Contrarily, if the generalization fits the established corpus of knowledge, it lends added credence to it.

It will be in order to elaborate at some length the points made above. The role of general propositions in enlarging or extending systematic knowledge may take various forms. The generalization can be instructively used to understand a certain region of phenomena or aspects of them that were not examined in these terms before. That is, a new observation inhered by the generalization may afford a basis for a better and deeper understanding of a phenomenon. Reviewed in terms of the new observation, the phenomenon under inquiry may reveal itself as one governed by a law applicable to a class of phenomena which, in turn, may be a unit of a still larger class of phenomena. In this manner, a new general conception helps us to appreciate the deeper linkages obtaining amongst several concrete phenomena or classes of phenomena. It is thus that a new set of facts is brought within the grasp of the established corpus of knowledge in a meaningful way. In consequence, the bounds of knowledge get enlarged not just quantitatively (i.e., more units covered under a conceptual scheme) but also qualitatively (i.e., in the depth of understanding and in bettering predictions). For example, a general conception of "relative deprivation" may serve as a useful means for examining the relation between rise in wages and productive efficiency of a worker. The problematic and paradoxical observation that a wage-rise does not always accompany productive efficiency (it may at times even hamper it) can be more aptly understood by means of this abstraction.

This behaviour of workers, i.e., low productivity despite wage-rise, then appears to be a corollary of reference group behaviour. What we have gained thus is a deeper understanding of a specific phenomenon. This, of course, is only one possibility. The matter will become more clear when the relation between research and theory is discussed.

Another way in which a general conception may enlarge systematic knowledge is by bringing to light the seeming inconsistencies in the existing corpus of knowledge and attempting to reconcile these. For example, the findings of studies on the effects of "authoritarian" and "democratic" classroom atmospheres on learning

are not seen to be consistent. A new general conception in the nature of a research finding, helps us to reconcile these findings by pointing that whatever the classroom atmosphere, the important things is the extent to which the teacher behaves according tot he expectations of the students.

Yet another way in which a new general conception extends knowledge is by pointing out or identifying certain gaps in the existing corpus of knowledge and attempting to bridge these up. These gaps have a reference to the failure or inability of a conceptual scheme or theory to explain and account for certain aspects of a phenomenon to which it should expectedly or principally apply. The theory that "uniformities of social behaviour represent conformity to norms" betrays such a gap since it cannot explain the existence of certain social regularities which are not culturally mandated. This inadequacy of the theory represents a gap in knowledge. The role of a new general conception may be to bridge this gap and to the extent this is successfully done, knowledge gets extended.

The general conception or finding emanating from research may have another impact on the established body of knowledge, i.e., correcting the errors in it. If the new general proposition derived through manipulation of things, concepts or symbols is discordant with what one would expect logically on the basis of the existing body of knowledge or theory, the only alternative (since the empirical observations and corpus of knowledge cannot for long stay divorced from and immune to each other in mutual indifference) is to effect corrections or modifications in the original theory in the light of the new evidence. In case the existing scheme of propositions has no place whatsoever for the new observations, i.e., the situation is such that either the observations are incorrect or the theoretic system (assuming, of course, that the observations are correctly derived) then scientific obligation demands that the prevailing system of propositions, should be revisited and revised. Proving that something considered to be true thus far is in fact wrong, is also an important contribution to knowledge. We shall have an occasion to illustrate this point later.

That the new statements of generality emanating as research outcomes serve as scales for verification of the existing system of knowledge is only repeating the obvious. As has been noted, research involves testing to find out whether the empirical observations presented as general statements are in accord with the predictions that may be made on the basis of the existing body of knowledge. If the observations are in accord with these predictions, the corpus of knowledge which afforded these, gains greater strength and confirmation. If such is not the case, the system needs to be revised or even rejected. An important aspect of scientific activity is verification of conclusions which have found a place in the established system of knowledge. This involves frequentative scrutiny. The use an analogy, the inspector is duty-bound to examine before each performance, the ropes on which the trapeze artists in a circus perform their perilous tricks. The ropes should have indubitable reliability. Everything depends on it. So too, in science, the established facts must be really established, for they are the rallying points and the bases for the abstract superstructure. Should clarification prove them to be unreliable or false, like the ropes on our circus sample, they should be cut down.

**(d) Knowledge may be used for Construction of a Theory or Practice of an Art**

The extended, corrected and verified knowledge may be put to two possible uses:

- (a) Theoretical, (b) Practical.

Knowledge thus acquired, may be used for constructing theoretic models. In other words, knowledge may be organized into propositions and these propositions may then be meaningfully articulated to form a more abstract conceptual system affording estimations about a class of things or phenomena governed by a specified set of conditions. Such use of knowledge is often labelled as theory-oriented, and the activities of a scientist who seeks knowledge for the sake of building theories of 'non-utilitarian' import are often known as 'pure', 'basic' or 'theoretical' research. Knowledge for

the sake of knowledge, i.e., only for the satisfaction of 'knowing' is the attitude that underlies scientific activity of this order.

Knowledge, on the other hand, may be used as a means to some practical or utilitarian goal and not constructed just as an end in itself. For example, the knowledge about people in respect of certain things they need, like and aspire for, may be used to work towards their welfare, ameliorating their practical problems or even exploiting them for certain personal or 'ideological' ends.

Science is a double headed weapon. It can do immense good but it can be made to serve sinister ends too. It has powerful muscles but no personality. It cannot lead, but can only serve. It has a sharp eye for methods and tools but is blind to ends and values.

The researches which seek knowledge mainly for the sake of serving some practical ends are often called 'applied', 'action-oriented', or 'practice-oriented.'

'Social Research' is again a broad term having a reference to different kinds of scientific inquiries conducted in the field of social sciences and to some extent, the behavioural sciences (the distinction between social sciences and behavioural sciences is itself not very clear). Sociological research, for example, which may be an especial concern for quite, a few may be considered a part and parcel of the general category that is designated as 'social research'. It may be said that, all sociological research is 'social research' but not all 'social research' would qualify as sociological research. 'Social Research' would refer to a large class of researches while sociological research may be considered a sub-class within it. In practice, i.e., during conduction, of a study, it hardly makes much difference how one labels the study. What matters is the scientific character of one's procedures and how successfully one can solve the difficulty (theoretical or practical) that originated the study. Whether one calls a study 'A sociological study of old people' or just plain, 'A study of the old people', it does not make much difference as long as it is a scientific study of the old people. It is not the label that by itself determines a researcher's procedures, insights, emphases and his findings, it is rather his training and capabilities, which does it. A

rigid division of researches on the basis of the traditional division of sciences is not only not practicable but also not quite desirable. It is well to take heed of Francis Bacon's reminder that "The divisions of sciences are not like the different lines that meet in one angle but rather like the branches of trees that join in one trunk." Lavoisier, in 1773, stated in a memorandum to the French convention; all forms of knowledge are interwoven in one great tapestry and we are assured of one ultimate pattern and design because there is a unity behind all knowledge.

In modern times, research is often a corporate affair in as much as the complex techniques of collecting and processing data require interdisciplinary co-operation. The scholars belonging to different fields of specialised studies applying different forms and techniques work together and pool their knowledge and insights at one place to solve some common problem which initiated research.

As far as the social or behavioural sciences differ from the natural or physical sciences in quite significant ways and also, in as much as they share among themselves certain common problems of control, measurement, quantitative analysis etc. with more or less equal intensity (so much so that at one time 'scientific study' of social phenomena was thought of as impossible), it would not be improper to use the term 'soci'al research' to embrace all scientific inquiries within the large field of social-behavioural sciences.

It will do the subject no harm, if a broad definition of research is adopted as what is generally known as 'social research'. By this token, social research may be defined as "a method of studying, analysing and conceptualizing social life in order to extend, modify, correct or verify knowledge, whether that knowledge aids in construction of a theory or in practice of an art." It is seen that social research is nothing more than the application of scientific procedures of manipulation (purposeful control), of analysis and of synthesis at a higher level of generality, to the social-human phenomena with a view to test, modify and enlarge systematic knowledge about social facts and social life, generally. Thus, social research has a reference to an investigation focused on social phenomena, which

aims to discover new facts about social reality or verify old ones, to analyse their sequence, interrelationships, causal connections and natural laws governing them by means of logical and systematized methods. It is clear that social research does not aim at finding the ultimate truths. Rather, it aims to understand and clarify the behaviour of man, the social world he lives in, the relationships he maintains, the influences which are exerted on him and the effects these have upon him and subsequently, upon the social institutions of which he is a part and through which social behaviour is mediated.

There is a tendency to equate research with a particular method of investigation calls for comment. Such tendency stems from a misconception of the scientific method, resulting in the setting up of a criterion which ignores its many important contributions to knowledge. It hardly needs to be over-emphasized that a "a study is scientific when its data are subjected to a logical analysis resulting in the development of a theory whether those data are secured by experiment, by statistics or common sense.

The fact that experiment cannot be performed in a particular case should not negate the possibility of scientific study. The term 'experiment' is sometimes used in a more restricted sense, to apply only to the situations in which objects or events involved can be deliberately manipulated by the investigator. This amounts to assume that manipulation is just about the only method of control. As we said earlier, an astronomer cannot manipulate the stars and the planets and yet he can conduct controlled enquiries into their relationship because he knows the values of the variables. So also can the social-scientist study the miniature social systems in a controlled manner if he is able to determine the significant properties of these systems (i.e., groups).

The historic reason, why manipulation is sometimes confused with control is that at one time manipulation was necessary to reduce the number of variables to just two, thus making them amenable to mathematical treatment. The development of methods of 'multi-variety analysis' has removed the necessity of manipulation and laboratory. Though control is not synonymous with manipulation,

some scientists, as Ackoff tells us, consider it useful to make a distinction between the general class of controlled inquiries and special class of inquiries in which control is effected through manipulation. This special class they call 'experimentation' while the general class is designated as 'research'.

This practice has an unfortunate consequence of according to the non-manipulative inquiries, a status lower than the manipulative ones. In fact, the emphasis ought not to be on manipulation but on control, where it belongs. For our purposes, research may also involve experimentation.

## PURE AND APPLIED RESEARCH

Research may be motivated by a desire to understand or for the sake of knowing (often called 'pure' or 'basic' research). The division between the two, i.e., 'applied' research and 'pure' research, is a matter of degree, it may be said that the 'pure' researcher derives greatest satisfaction from increasing his knowledge in a field of inquiry where many questions remain unanswered. To him, the challenge of not knowing is paramount. If he can solve the problem, he is satisfied; the results may or may not have any practical use. The 'pure' scientist would probably argue that knowledge itself is in the last analysis, always of some practical use. To him, scientific enquiry is noble in itself; it is its own reward. To keep digging away at the layers of intellectual questions is a challenge enough for a 'pure' scientist. To him "knowledge is the highest good, truth is the supreme value; all the rest is secondary and subordinate." The 'applied' scientist, however, is apt to see his research in a practical context. He would be inclined to define a problem as one with reference to which some action could be taken to improve the matters.

The applied scientist is thus much more lively to be working within a certain set of values and norms to which he feels committed. Sociologists for instance, study a host of problems, viz., the problem of juvenile delinquency, of old people, of gangs, etc. The 'pure' sociologist would study them just to find out the 'what' and 'how' of these, whereas the 'applied' sociologist would study these mainly

with a view to finding out how best to go about arresting delinquency or helping old people. The applied scientist, of course, is working on the same data as the pure scientist, but while the latter would perhaps stop short of advocating changes, the applied scientist is prepared to step in and propose the measures to ameliorate the situations. But once, he starts to indicate what ought to be done, the applied scientist is stepping outside the structure of scientific investigation and is becoming a moral human being devoid of (scientific) neutrality.

As is suggested earlier, the controversy between 'pure' and 'applied' science need not be made much of, because the two are not mutually exclusive and there is a continual interplay between them, each contributing to the other in many ways. Says Oppenheimer, "the great testimony of history shows how often in fact the development of science has emerged in response to technical or even economic needs, and how in the economy of social effort, science, even the most abstract and recondite kind, pays for it self again and again in providing the basis for radically new technical developments."

Much is discussed in favour of both sides, viz. the 'pure' science and the 'applied' science. Huxley disliked the term 'applied' science heartily and pointed out the risk in over-emphasizing the utilitarian criteria for judging the worth of sciences. Francis Bacon, a staunch advocate of 'pure' uncontaminated science, said, "just as the vision of light itself is something more excellent and beautiful than its manifold use, so without doubt the contemplation of things as they are without superstition or imposture... is in itself a nobler thing than the whole harvest of inventions." It is an intolerant narrow-minded view which supposes that a science is to be judged only by its practical fruits and not also by illumination it affords.

Morris Cohen's perceptive statement summarizes the importance of pure research, i.e., the purely theoretical contribution of research. He observes, "purely theoretical contributions to astronomy and mathematics, by increasing the precision of navigation have saved more lives at sea than many possible (technological) improvements in carpentry of life-boats."

Parsons also argues "It is not a question of whether we try to live up to our social responsibility but of how? If we should put the overwhelming bulk of our resources... into immediately practical problems, it would do some good but... it would have to be at the expense of our greater usefulness, to the society in future. For it is only by systematic work on problems where the probable scientific significance has priority over any immediate possibility of application that the greatest and most rapid scientific advance can be made." He posts the importance of 'pure' research directed at solving abstract theoretical problems. To him the application is only a technological by-product of abstract relationship among concepts, which is potentially a generator of many more applications.

The history of science testifies to the fact that the practical value of science has been in direct proportion to the precision of scientific methods and that most 'theoretical' or 'pure' investigations have practical results of extraordinary magnitude. No great law of science has ever been discovered for its practical applications but instances abound of investigations apparently quite useless practically, which have led to most valuable results.

To quote Dixey, "natural knowledge pursued for its own sake without any direct view to future utility will often lead to results of most unexpected any direct view to future utility will often lead to results of most unexpected kind and of very highest practical. Thus, undue insistence on the practical utility of science is not historically justified and a hasty criticism of a certain scientific work being purely 'theoretical' is likely to be very unjust.

In contrast a view is held that "Science is for life, and not, life for science." R. S. Lynd's book, *Knowledge for what?* gave vehement expression to the feeling, "what good was knowledge if it did not have practical utility in terms of affording solutions to day-to-day problems?" Samuel Stouffer, the research-director of the well-known 'American Soldier Studies,' saw considerable merit in the insistence on the practical and its value for 'pure' research. He pointed out that practical application of a science enhances public recognition of the importance accorded to it and influences public

co-operation and financial support for its prosecution. Secondly applied research stimulates the improvement in the known tools and leads to the discovery of better ones. 'Pure' science benefits from techniques developed by practitioners in applied fields. He further points out that 'applied' research can speed up the process of building basic theories. The pressures to 'explain' or interpret surprising or unexpected practical findings, may lead to a reflection which organizes many such findings. The concept of 'relative' 'deprivation' for instance came to surface owing to the pressure to explain an anomalous or surprising observation in a practical situation, for example, the Negro soldiers who wanted to be stationed in the northern American camps were less well adjusted to army life as compared to the Negro soldiers who were stationed in southern camp against their wish.

It has been discussed that the interplay between 'pure' research and 'applied' research. There are contributions to the other from each side. As Dewey aptly observes; "the method and conclusion of science do not remain penned within the science. Even those who convey of science as if it were a self-enclosed, self-actuated, independent and isolated entity cannot deny that it does not remain such in practical fact." Finally power of science appears to lie in the development of general principles which are applicable to many practical problems.

As is evident that 'pure' and 'applied' researches do not constitute an 'either-or' dichotomy. Historically, knowledge for its own sake and knowledge for practical application have both characterized the scientific enterprise. "... it has always been a tradition in science guiding the activities of many generations that science should be applied to practical purposes and the practical application should be a check on the validity of results and a justification for the efforts." The dual emphasis is meant to provide for the development of a body of general principles about human (nature) behaviour and relationships. Secondly, because of its social orientation, it is expected to provide practical guidance in solution of social problems.

## QUALITIES OF A RESEARCHER

A question quite often asked is "What are the qualities of a researcher?" To this question, a simple answer would be that being a scientist, the researcher should be firmly committed to the 'articles of faith' which underlie scientific method. This only amounts to saying that a researcher should be a man of science in the true sense of the term.

Sir Michael Foster's statement in identifying the distinctive qualities of a scientist cannot perhaps be bettered. He has specified three qualities of a true scientist:

(1) Above all other things, his nature must be one that vibrates in union with that of which he is in search, the seeker after truth must himself be truthful, truthful with truthfulness of nature; which is far more imperious, far more exacting than that which man sometimes calls truthfulness. Truthfulness corresponds to the desire for accuracy of observation and precision of statement. First, to make sure of facts is a fundamental precept in science; but this is not an easy matter. The difficulty here may be due to the untrained eye, which sees only that which it has the power of seeing, sometimes little indeed. It may be due to preconceptions which often make men see what is not to be seen. It also may be due to lack of discipline in the method of science. The unscientific man is often content within 'approximately', 'nearly' and so forth but nature never is. It is not her way to call the same two things which differ, however minutely. She resents the conduct of men who treat such differences in any other way than she treats them herself.

(2) The man of science must be of alert mind. Nature is ever making signs to us, she is ever whispering to us the beginnings of her secrets. The scientific man must be ever on the watch, ready at once to lay hold on nature's hint, however small, to listen to her whisper, however low. Receptivity to the hints and gestures of nature is something that has to be cultivated slowly and patiently. It is not given to the ignorant and the common place to see the unusual behind the routine. It demands a systematic immersion into the subject

of concern to be able to catch the slightest hint that might give birth to significant research problems.

As Cohen and Negel so rightly point out, "the ability to perceive in some'brute experience, the occasion for a (research) problem is not a common talent among men. It is a mark of scientific genius to be sensitive to difficulties where less gifted people pass by untroubled by doubt."

(3) Scientific enquiry, though preeminently an intellectual effort, needs a moral quality of courage; not so much the courage which helps a man to face a sudden difficulty as the courage of steadfast endurance. The prosecution of science is a thorny affair. There are times when the scientist feels defeated and lost in wilderness. This is when he needs a supreme courage of conviction. He must learn to endure hardship intellectually. Darwin said, "It's dogged that does it."

The sacrifice demanded of a scientist at the altar of truth requires no less a courage than that exemplified by Von Siebold and associates who heroically swallowed bladder worms to verify the truth of their 'theory'. It takes qualities of courage to be able to stand by one's conclusions or scientific conviction at the risk of social disapproval.

The quality of cautiousness might be the fourth quality of a researcher. As Huxley puts it "the assertion that outstrips the evidence is not only a blunder but a crime." The researcher should habituate himself to withholding any judgement when data are obviously incomplete. According to W. K. Brooks, "the hardest intellectual virtue is philosophic doubt; the mental vice to which we are most prone is our tendency to believe that lack of evidence for an opinion is a reason for believing something else.... suspended judgement is the triumph of intellectual discipline."

Cautiousness then, is the essence of science. One form of cautiousness, most difficult of attainment and yet indispensable, is distrust of our personal bias in forming the judgements. Karl Pearson rightly remarks, "the scientific man has above all things to strive at

self-elimination in his judgement, to provide an argument which is as true for each individual mind as for his own".

## NATURE OF SOCIAL RESEARCH

The distinctive nature and character of social research derives in a significant measure from the real and supposed nature of the social phenomena which poses certain difficulties when it comes to application of the sophisticated scientific procedures characterizing the natural science to social phenomena.

It does not mean, of course, that social sciences are not sciences in any real sense. We may like to see what the typical limitations of social science research are:

In case of the softer social sciences so little spontaneous guidance is afforded by the subject matter than in some of the natural sciences which have a logic of their own unobtrusively pointing the way that substantive research often yields place to repetitious discussion of methodology. A natural scientist need not worry about his laboratory experiment getting vitiated by his mood or by declaration of a foreign policy and other social events while all these and many more such factors need to be carefully controlled lest a social scientist should foul up his work. Also, certain properties of the subject-matter that social scientists deal with give rise to special problems. The results of social science inquiry are statistical, that is, presented in probability terms. They are never strictly categorical and clear cut. A certain new technique of advertising may work better than the conventional method with a certain proportion of manufacturers while, the conventional method would appear to suit the rest. In other words, the differerices between two or more categories within a social system may be so small that nothing can be said conclusively on the basis of comparison.

Besides, more than one important variable is generally involved in social science problem. Quite so often, it is nearly impossible to segregate or disentangle the different variables to ascertain their effects individually. This difficulty is nearly unsurmountable when these different variables operate jointly and also are not amenable

to experimentation. For example, when people are less educated they usually have lower income, too. Hence, it is difficult to determine whether less educated people are less mobile than other people because of their education, because of their income operating together, or because of both.

Furthermore, the researcher himself, being a human being, a member of groups, a buyer etc., frequently affects the subject-matter and in effect changes the whole situation.

If contrasted to natural sciences the social sciences can barely construct a complete system. A physicist can discuss and set up equations for the entire system in which electrons flow in a circuit. But interactions among human beings cannot be so described for the human 'systems' get punctured so often and so easily that predicting a long-term sequence of events becomes well nigh impossible since a number of new influences enter with each interchange among human beings; the system is never really closed.

Quite too often, the social scientists are criticised for labouring hard to contribute small bits of knowledge. In all fairness to the critics it may be said that they are not far wrong when they lead a criticism of this order. Conceivably, two factors can explain this poor out-put of social scientists. There is no denying the fact that many a serious problem of the world has barely been scratched. What causes war ? What will ensure peace ? Why so much of dehumanization ? Social sciences may, of course, be said to have made some gains in certain fields especially in the economic discipline, but many a huge human problem still remains untouched.

However, social scientists used to learn much about our social world in a very simple way. Most social questions can be answered very easily. Why people in slums suffer from ailments ? Because they are poor. Why don't many women seek legal redressal of their genuine grievances ? Because they feel it is unbecoming of women as a class to do so; on and on. Such and many other questions can be answered quite well (we feel) by consulting our own experience, by relying on our habits of thinking or by asking other

people. These questions are nowise trivial but most of these have no single clear answer. In each case, some sole factor is held responsible for what happens. But in the natural sciences, it is a major task-to unravel even such problems.

The behaviour of human beings is affected by diverse influences such as environmental, temporal, biological, psychological and socio-cultural, all of them affecting it contemporaneously. The complexity of human or social data may be largely attributed to this. It is difficult for an observer to see the underlying uniformities in the profuse diversity of human behaviour which is in a sense unique for each person. Hence, it is a formidable task for a scientist to discover an order or principle which would apply to all men or the bewilderingly complex human data.

In social sciences, the laboratory is the society and the objects are conscious and active human beings. "The observer and observed both being similar become so confused that an objective approach is really difficult to make. Moreover, except in a totalitarian society, a controlled experiment in the laboratory of society with free men as objects... *sine qua non* of an empirical science, is well nigh impossible in social science, generally.

At this stage, it may be well to lay emphasis on the distinctive features of Social Sciences. For instance, the complexity of social data is not so well founded. In the midst of the apparent chaos, there is indeed some pattern. If social life were so utterly complex, it would be unlivable. All social interaction is based upon expectation of behaviour, it may be an interaction between thousands of people in highly complex groups or an interaction in small cohesive groups. This means that a reasonable prediction about people's behaviour is possible, may it is an important aspect of social life. By complexity, one must realize that complexity is a relative term. Social phenomena are complex to us, because our knowledge of them is inadequate and our tools of study have developed a little beyond infancy.

It is pointed out above that as contrasted to the physical sciences, the social sciences lack the power of exact prediction. This is attributed to the 'erratic, idiosyncratic and irregular' nature of human behaviour.

The case for unpredictability of social behaviour is again, not so well-founded. While individual behaviour may be unpredictable, one can predict with quite a high degree of accuracy the behaviour of a whole group (on the basis of knowledge of the pattern). Lundberg has rightly pointed out that the low predictive potential in social sciences is due mainly to our limited knowledge of relevant variables operative in the groups. "As our knowledge of the variables increases and we are able to judge the effect of various variables involved, it will be possible for us to predict social events with much greater accuracy."

Whereas the physical phenomena may be known directly through senses, the social phenomena are known only symbolically through words or terms referring to such phenomena, e.g., tradition, custom, values and the whole range of the subjective world. This makes verification of conclusions very difficult.

It may be pointed out in this connection that there has come about a standardization of concepts connoting social facts, and also that techniques have been developed to measure most of the so-called subjective items in objective terms, e.g., anthropometric or sociometric measures.

Lundberg feels that most of the subject-matter of social science is qualitative and does not admit of quantitative measurement. This contention leaves some ground for criticism for, qualitative and quantitative measurements are only different stages in the growth of a science and it is not as though some data are by nature quantitative and others qualitative. As science develops, what were previously thought of as qualitative data may be transformed as quantitative ones. Secondly, we should not forget that qualitative expressions and analyses have their own place of importance in a social inquiry.

It is argued that social phenomena compared to the physical ones are characterized by greater heterogeneity. Even if we accept this, it is possible by adequate stratification or classification effected in terms of certain traits or properties, to ensure a fairly high degree of intern homogeneity within each stratum or class. Thus, social research may reach conclusions of broad applicability.

It is conceded that most physical sciences also known as exact sciences, allow for controlled laboratory experiments, hence their exact ness. The social sciences suffer from this handicap, although to a limited extent laboratory experiments are possible here too. As social science develop, a number of human problems may hopefully be brought within the reach of laboratory experiments.

One of the characteristics of the social phenomena is that the cause and effect (better still, the producer and product) are difficult to be segregated clearly. In social sciences, it does not many a time, make sense to ask which is the cause and which is the effect (e.g., poverty and lack of skill). It is obvious that unless we realize this fact, we shall be asking wrong questions and finding wrong answers.

Social data typically pose certain problems when it comes to these being treated by the highly developed quantitative methods of the physical sciences. It should be now clear also that quite a few critiques of the stature of social science research do not hold much water. Atleast the difficulties are not insurmountable. The difficulties which appear to preclude the possibility of a 'science of society' derive from our underdeveloped techniques and methodology of study and our consequent unfamiliarity with data rather than from the inherent differences between the data related to the two types of sciences.

It is a fact that social sciences in their present state of development are far behind the physical sciences. Says R.K-Merton, "We social scientists, happen to live at a time when some of the physical science have achieved comparatively great precision of theory and an abundance of technical by-product... many social scientists take this as a standard of self appraisal... they want to compare biceps with their bigger brothers. But this is to ignore the distinctive forehistory of each: between 20th century physics and 20th century sociology stand billions of man hours of sustained disciplined and cumulative research."

Merton advises that social scientists not to express despair and harbour doubt whether a science of society is really possible,

but, with the present limitations in view, "develop special theories applicable to limited range of data" and slowly build their way up toward more general theories broader applicability

### **UTILITY OF SOCIAL RESEARCH**

To the question "what use is of social research?" one may reply "of what use is a new-born child?" in the manner of Benjamin Franklin who replied thus, when asked about the utility of his findings about the relationship between thunder-clouds and electricity. This means that new knowledge like the new-born baby, holds great potential of growth and maturity. Also like the new-born child, it gives us pleasure. It gives us satisfaction of knowing the unknown. This points to a value that the scientist is committed to, i.e., the self-justifying goodness of 'new knowledge' about anything big or small. "Social research is persistently opening our eyes to the social reality, simplifying the mysterious within the seemingly commonplace in social life and shattering its garments of make-believe by which pious hands have hidden their uglier features. The obvious function of research is to add new hidden their uglier features. The obvious function of research is to add new knowledge to its existing store, but its power of cleansing our minds of cliches and removing the rubbish of inapplicable theory are equal notable. Scientific research is a cumulative process... it is also a rejective process, especially in social sciences... understanding can be (advanced) no only by gains in knowledge but also by discarding outworn assumptions."

A social researcher is interested in the discovery and interpretation of social processes, patterns of behaviour, similarities and dissimilarities that apply to typical social phenomena and social systems, generally. That is, the social researcher is concerned with types and classes of social situations persons or groups of which the unit he is studying at the time, is a specimen or an instance. His facts are selected and related according to their intrinsic nature and their susceptibility to organization into a logical system. This search for knowledge has a definite relation to people's basic needs and welfare. The social scientist assumes that all knowledge is potentially useful in the end. It must be remembered, however that science

and society have a two way relationship. There is a give and, take between science and social conditions. Science helps to create social conditions; social conditions recharge the accumulators of science.

In a general way, some of the direct practical benefits and theoretical implications of social research may be listed as follows:

(1) Social research has a crucial role to play in guiding social planning. Adequate social planning depends for its success on a systematic knowledge of the social resources and liabilities, of the people and their culture, of their similarities and differences, of organizations and operative controls, of their needs, hopes, aspirations and problems. Any effort at social planning is bound to fail if it is based on fictitious assumptions of planners in relation to what the consumers of planning need, what their problems are, what they want remedied, and what kind of a system they want as an emergent product of planning.

Social planning, or for that matter any planning, requires a store of reliable, factual knowledge on the basis of which a blueprint may be designed and the difficulties in its implementation anticipated and guarded against. Nor is it all; such a foundation of scientifically gathered knowledge affords a basis for evaluating the net gains of planning for the social system in question. Social research is of immense help in securing such knowledge. It happens so often, that the overly zealous practical men with a programmatic orientation consider social research an unnecessary expense only to realize subsequently that the factual data would have helped them avoid the vast wasteful expenditure of money, time and energy occasioned by the failure of their designs on the plane of practice. Social research is generally worth much more than the costs incurred over it.

(2) Since knowledge is a power, social research, by affording first-hand knowledge about the organization and working of society and its institutions, gives us a greater power of control over the social phenomena and action. Thus, social research may be visualized as having practical implications for formal and informal types of leadership patterns of influence and reform in different spheres of society.

(3) It is rightly said that knowledge is enlightenment. It dispels the thrust of outworn assumptions, superstitions and stereotypes. Social research, thus, may be expected atleast to afford a more solid basis for people to hold whatever opinions they do. Some authors have claimed that social researches may have the effect of promoting better understanding and social cohesion, since it brings to light the underlying oneness in the midst of a bewildering variety or diversity of human societies. But this is claiming too much for oncside and ignoring the other possibility; social research may also unravel diversity in the midst of apparent unity.

(4) It is clear that social research has direct implications for social welfare. By virtue of the deeper understanding of the causal nexus underlying various social 'maladies', social research provides a secure basis for effective remedial measures. Social researchers analyse the problem in the 'total context' (this is desirable) and as such are in a better position to identify social structural anomalies and ambivalences that get reflected in the form of these problems and hence, structural changes would be necessitated. The 'remedies' suggested by research are thus deep going. They hit where they must. Many of the laymen reformer's 'remedies' create other new problems or 'side effects'. Scientific social research thus provides sound guidelines for appropriate measures of welfare or reform. It is no accident that a large portion of legislation and reformative measures owes its origin to reports of social surveys.

(5) A researcher is loaded with the responsibility of ascertaining some order among facts. Thus, research affords a considerably sound basis for prediction. Despite the admittedly low predictive potential of social research, reasonably reliable predictions, perhaps 'culture-bound' or 'context-bound', can be made. These have the effect of setting our efforts a social planning and control on a sounder footing. The success of planning for social development depends to a great extent on our intimate knowledge of our own society as also of other societies. Thus, social research has the effect of initiating and guiding social growth on proper line and towards the cherished goals.

(6) Every scientist is obliged to effect constant improvements in the tools and techniques of his trade, i.e., research. The social researcher, in so far as he has to work in reference to different spatial-temporal contexts, each challenging his attack, is constantly faced with the need to improve upon his tools or if need be, to fashion new tools to match his skills with the task prompted by the exigencies of the situation. Samuel Stouffer and his associates working on the adjustment problem in the context of racial prejudice to cite only one instance out of the many had to effect modifications in the prevailing techniques of research and when occasion demanded, to invent new ones to take the best out of the situation.

In developmental parlance, the major possibilities of utilizing social science research may be identified as under :

(a) Social research may hardly afford valuable background data to be capitalized by social planners for assessing the existing state of affairs; particularly the magnitude, complexity and ramifications of the problem they are expected to grapple with; the critical may be illuminated by analytical studies. The observed and hidden dimensions of the problem thrown up by such studies may be expected to proffer certain measure of foresight to planners to deal with the problem effectively.

(b) Such social science exercises may provide a basis for testing the validity of certain assumptions that our planners inevitably have to make in laying down their short term and long term goals. These researches conceivably, may help the planner to anticipate the consequences and cost of alternative strategies which may be pressed into operation for achieving the settled for goals.

(c) Social science researches may bring-to sharp focus the varied influences and factors that contribute to the failure of certain projects. Hence the policy planners may stand forewarned about these.

(d) If social science research findings become a part of public knowledge a general awareness about the situation and challenges, as also the desired policy to meet them squarely may result. This

would prepare people for accepting a particular policy and for exerting popular pressure for reformulating or amending current policies, or rejecting them.

## SOCIAL RESEARCH IN INDIA AND ITS IMPORTANCE

The field of social research is as broad as human behaviour and its systemic manifestations. A realization of what this means should make it clear to us that the store of potential problems for social research is almost inexhaustible. The state of social sciences with the exception perhaps of economics, is characterized by researches being conducted with a view to propounding special theories of limited applicability. These sciences are not yet prepared to ascend to the level of higher generalities. The long hard climb up would involve, firstly, a patient building up of what Merton terms, the middle range theories, applicable to not too broad and not too specific aspects of human phenomena and secondly, their synthesis in more general conceptual schemes. The social sciences, as of now, are not far up from the bottom of the ladder.

In this continuing process of social research, quite a few specialized theories have emerged. The emphases or concentration on areas of study and avenues of theorizing has kept shifting from time to time. Unfortunately, very few areas (if at all) have been subjected to an intensive attack in terms of sustained attempts at replication or verification of theories, with the result that many of the prevailing theories might perhaps be on the level of plausibility. It is also possible that some of them might be 'context-bound' or 'culture-bound' (i.e., valid only in a particular spatio-temporal context or a particular culture). What is needed, therefore, is the testing of such 'theories' on the anvil of different concrete socio-cultural contexts. Needless to say, the theories to be representative must fit different social situations of a given class. Hence researches are needed to probe whether the theories hold good, irrespective of contexts.

It is well known that some anthropological studies, like M. Mead's study of Samoa, conducted in simple 'primitive' societies

sounded the death-knell for some earlier sociological and psychological theories. Just as the 'primitive' societies serve as testing points for certain theories of human behaviour, so also may the developing countries like India serve as test bases for the existing theories of (just to mention, two) social organization and change. It is not difficult to appreciate that replication of existing ideas in a new context, besides testing the theory developed elsewhere, enriches our understanding of what the theory principally applies to. The developing countries are eminently suited as an anvil for the testing of theories or hypotheses of social change; for instance, of modernization of development, because it is here that one can vividly see change as a live process. A study of 'modernization' in the Indian context which may be considered a unique magnetic field of social forces pulling in opposite directions, is rightly expected to be quite revealing and enlightening besides its significant feed-back value for theory.

Questions are asked: What are the tangible indices or referents of modernization? How can revival of certain traditional patterns in India be reconciled with the process of modernization? What would be the phases on the continuum of modernization? How can these be rendered in operational terms? just to mention a few, are questions that social research may be expected to answer. To expect the least, these answers can help clarification of the concept which is no small a contribution to theory building. It should be clear now that research designed to test hypotheses deduced from social theories developed in different socio-cultural contexts, besides testing and clarifying the theories by unfolding the ramifications of their operation, also help to define them by pointing out their limitations and thus add to their representativeness. Social researches conducted in different socio-cultural contexts thus go a long way in providing rich insights whereby different concrete phenomena may be grouped under a concept, and by enlarging the bounds of the concepts, i.e., the building blocks of theories, help to enlarge the scope of the theories. Thus, the process of 'Sanskritization' may be seen as an instance of reference group behaviour and as such may become amenable to scrutiny in terms of the orientation afforded by the

theory of reference group behaviour. Needless to say that any specific phenomenon in a particular socio-cultural context which may legitimately be considered a concrete referent of a concept but which does not conform to the expectations (predictions) based on the theory of which the concept is a part, introduces a problem. The solution of this problem by research has definite consequences for the theory in terms of its possible revision or clarification or even rejection. But this has to be the function of research conducted anywhere as we shall see more clearly when we examine the place of empirical research vis-a-vis theory. Why then, do we have to essay highlighting the importance of researches conducted in diverse socio-cultural settings, especially those preferred by the developing societies?

Outcomes of researches conducted in particular socio-cultural settings, i.e., the generalizations or theories, cannot be expected to apply automatically to other socio-cultural contexts since typically, these contexts differ from one another in important respects. The natural sciences, in this respect are in a more fortunate position attributable in turn to the homogeneity of their subject matter. (This is not to deny that this alleged homogeneity could in part be a result of scientific insight and depth of understanding on the part of those who deal with the subject matter.)

Due to heterogeneity of the subject matter of social sciences may be attributed the reason why the social scientist is skeptical of (and, it is a desirable quality of the scientific mind) conclusions and theories based on researches conducted in particular societies. A virtual transplant of research products, i.e., theories, on to any other socio-cultural context uncritically is rightly improper because the native soils are different. Spatio-temporal contexts and cultures operate in too subtle and imperceptible ways in influencing human behaviour. This warrants acceptance of these products of research only after the screws of scientific test have been properly applied to them. It is here that researches in diverse socio-cultural settings have their own importance. The developing countries have a very special status of import in this respect, precisely because these

typify the conflicting social forces; some of which work for continuity of certain patterns and others against it a conflict between tradition and modernity or industrialism, for instance. Such societies are characterized, on the one hand, by 'norms' that are pressing and striving for social acceptance or institutionalization, and 'norms' which are on their way out but not quite gone, on the other. Such societies are not at rest with themselves; no society ever is; but unrest here is undeniably critical. In them, the social processes can be seen in their highly accentuated form; their problems are serious. So not only as testing points for theories but also in view of sheer richness, profuseness, challenge-ladenness of their social realities, such societies have a pride of place as areas for fruitful conduct of research.

The practical value of social researches for such societies in terms of social policy formulation and planning, etc. shall be briefly illustrated with reference to India. For the sake of convenience, we shall confine our discussion on the utility of social research of India to the period after 1947, for, the attainment of independence heralded a new era of consciousness and development, of progressive policy formulation and legislation, and of multitudinous changes and attendant problems. This was when the new born democratic nation started to take its first faltering steps towards its distant cherished goal of welfare through socialism.

Social planning is prerequisite. It may be the planning for development material, social and cultural, in such a manner that the weaker and more vulnerable sections of the population, in particular, would benefit from its fruits and smoothly enter the mainstream of national consciousness.

With this in view, the Planning Commission, constituted of outstanding persons representing major fields of expertise, formulated development plans which were enthusiastically implemented.

The results, however, are not always what the planners would have liked them to be; lot of divergences were often evidenced between the overt aims of policy and its consequences in practice;

there was considerable wastage of national funds, energy and talents. The evaluation reports brought into sharp focus the need for systematic social survey/researches in areas where plans were proposed to be implemented. This need was recognized by planners themselves but the evaluation reports afforded a vivid demonstration of how social surveys or researches could have avoided much of waste. It was not long, before the planners realized that if planning was to be meaningful to people for whom it was intended, if it was to be realistic in a particular context of needs and aspirations, if the programme of its implementation was to be a people's programme and not one foisted upon them, then it should be based on complete and reliable data on the people who are to be the benefactors of planning, their need structure (priorities), their values, attitudes, their strengths and weaknesses. That 'social' surveys/researches are as important as the 'physical' surveys of the area in that the success of programmes depended, in the ultimate analysis, on people's acceptance and participation, was a realization which gave impetus to the initial setting up of Research Programmes Committee in 1953 under the aegis of the Planning Commission of India and its eventual culmination in the establishment of the Indian Council of Social Sciences Research (ICSSR) in 1969.

The community development programme, an integral part of the Five Year Plans, was perhaps the most important single constructive activity "so big in content and so revolutionary in design," undertaken by the national government to bring about a rounded development of rural India in a planned democratic manner! Persistent difficulties in the implementation of these programmes in different parts of the country have brought about a clear realization that planning for human material presupposes a thorough knowledge of this complex material. Failure to appreciate this was bound to decrease the fruits of planning for a community. Each community is in a sense unique, hence no standard designs of a programme can be rewardingly applied to a community, neglecting its peculiarities in terms of human needs, cultural values and the unique intermeshing of forces operating therein. Thus, the systematic studies of communities came to be

reckoned as a prerequisite to any successful developmental programme.

An unimaginative application of the C.D. formula to the tribal communities has time and again illustrated how important it is that a community should be developed in terms of its own values and goals. It has been pointed out by many an official that our ideas of development do not augur well with those of the tribals. What we deem necessary for a 'civilized existence' (so-called) has no place in their scheme of thinking and mosaic of values. Hence, any scheme for their 'development' unrelated to the view-point and the philosophy of the tribals will not only be an unwarranted interference into their affairs but may bring them more harm than good. Hence the need for scientific studies conducted to ascertain the whole kaleidoscope of the life pattern of tribals and of communities in general. It is gratifying to note that an important part of the role of tribal research institutes and anthropological surveys fulfills this need to some extent.

In a welfare state like India, in addition to the general social services like education, health, etc. and the specialized social services intended for the benefit of the weaker and vulnerable sections of the population, like, women, children and the physically, mentally and socially handicapped, the state is in the process of passing progressive, protective legislation. The social welfare services need professionally competent personnel serving the cause of the handicapped of various kinds. Such services in order to be beneficial and fruitful have to be based on reliable information about people whose welfare is being contemplated. Welfare in the true sense implies giving just the right kind and degree of support to the purported beneficiaries in view of their problems so that the afflicted will acquire enough strength, capability and motivation to help themselves grow. Such help or support that will not corrupt a person or a class of persons but will ennable them, can only be administered on the basis of reliable data afforded by scientific social studies. It hardly needs to be emphasized that legislation which is a means to amelioration, must be based on a reliable assessment of the needs and feelings of the people and not on mere assumptions.

An up and coming industrial nation, India is faced with certain characteristic challenges' which need an urgent coming to grips with. The first and the most obvious social alteration of the country under industrialization has been in the productive organization itself. Compromises with traditional norms of conduct, their effect on workers' efficiency and the extent to which such compromises impede full social transformation are important questions concerning work relations in India. Industrial technology has imposed some minimum requirements on the organization of productive units. The bureaucratization of labour force an accompaniment of industrialization, presents various transitional problems not only for factory recruits but also for the elite groups. The need for subjective attitudes and objective incentives favourable to novel jobs and conditions of employment is implicit in the requirement of new work-relations and the new forms of productive organizations. Various obstacles in the way of fuller commitment of persons to the norms of industrial work are evident. Industrialization involves gross changes in the occupational structures mooted by the economic sector and finer changes necessitated by demand for specialized skills. A vital economic problem of India is the shortage of capital; the budgetary behaviour of people also needs to be taken into active account.

The population explosion and the rapid rate of urbanization are evidently connected with industrialization in India. infant mortality has gone down; there is in evidence a high cityward migration; age composition has changed and with it has emerged the problem of dependency. All this has serious implications for the varied spheres of social life. In India, cities seem to have grown at a rate surpassing both expansion of employment opportunities and of public services or amenities.

Technological modernization in India appears to be linked with profound changes in the family and kinship organization. A common symptomatic feature of industrialism appears to be the disintegration of the family. Industrialization has given rise to many a problem of community organization, e.g., problems of residence, transportation, and congestion. Slums have appeared, constituting

one major environmental source of a number of problems of social control. Crucial questions of social strategy are now involved in formulating an educational policy in view of industrial demands. The correlation between education supply and man-power demand is hardly perfect. The mass media of communication supply the technical basis for promotion of a national public participation in political events and policy-making. Political participation seems to have been organized around unions or other occupational groups for a variety of interests and loyalties. Industrialization seems to have influenced, in important ways, the religious organization and beliefs.

Industrialization has provided a new set of social positions and new criteria of placement and valuation. It has, therefore, resulted in the complication of the system of social stratification. This has been responsible in some measure for status ambiguities.

The above is a very brief outline of the major social implications of technological modernization or industrialization in India. The manifold and multi-faceted implications of industrialization for India are seriously spell out the need for scientific social research on a fairly large scale, to provide or a systematic understanding of nature of problems, a sound diagnosis and an effective treatment plan. In fact, the complexity of the situation (of which little is known) calls for some kind of an inter-disciplinary collaboration researchers from specialized disciplines economists, political scientists, sociologists, psychologists and anthropologists, etc., making for an effective multi-pronged attack on these social maladies, and for devising long term social policies of real consequence.

## **STATUS OF SOCIAL SCIENCE RESEARCH IN INDIA**

The role of social science research in the social and economic development of the country is being increasingly recognised in India. Besides the government, public and private organisations are providing impetus to researches in the social sciences. The notion that social science research can contribute variously to the social, cultural and economic development of the country has been gaining ground ever

since the Planning Commission after its creation realised the need for social science research in the relevant spheres of national development. The universities and other research institutions have been activated to work out and administer various schemes of research into social, economic and administrative problems germane to national development. The first five year plan provided for a sum of rupees fifty lakhs to further these objectives.

Among the activities of the ICSSR, promotion of social science research has a central place. These promotive activities consist in financial support for research to social scientists and institutes and award of fellowships. As many as 708 research projects were sanctioned by the ICSSR during the period commencing from 1969-70 to 1976-77. The development of the institutional base was mainly a responsibility of the university system. But subsequently, the needs for social science research outside university system was also recognised.

The University Grants Commission has stepped in a big way to promote social science research. Since the year 1975-76, the UGC has been implementing the teacher fellowship scheme. The UGC also has a scheme for award of scholarships and fellowships for the development of social science research in universities and colleges. There are certain foreign agencies like USEFI and Rockefeller Foundation which provide financial support. Big industrial organisations and business families have also provided assistance for social science research. Social sciences, however, have had a low priority in the matter of research funding. The allocation of funds for research as compared to the natural sciences is meagre.

Social science research according to S.C. Dube, "is not a substitute for national development; in many instances it would not even indicate a clear policy line. Social scientists are nowhere close to being social engineers.... The power of social science is exploratory and its practitioners are essentially analysts."

It may be critically evaluated that the present status of social science research in India, mostly lacks a national focus and it is not

relevant to the major issues of the day; very few right questions are being asked by social researchers and even when asked they are not pursued seriously. Outworn methodology is being used and there is little evidence of methodological innovation which will prove equal to the unique problems of the Indian context. With heavy reliance on concepts formulated in foreign soils and uncritical use of borrowed methodology. Dube laments, that "social scientists in India tend to work upon the trivial and count barren theoretical and methodological exercises as triumphs. Alternatively, we refuse to examine empirically the social reality and try to answer the contemporary questions from the ancient texts and traditions."

There are a few exceptions of good research work in the realm of social science but these have a narrow focus. The insights emanating from them are hardly subjected to systematic verification and hence the extent of their applicability to wider areas remains unknown. As a result of it, a macro-national perspective cannot emerge. Hence, while such studies do fill a gap, it is difficult to generalize from them.

In last few decades, financial assistance is being given on an increasing scale to support social science research. A considerable volume of data has been gathered through such funded projects. But, how far these data were utilized for planning ? Knowledgeable persons feel that planning was by and large done in total indifference of these data. Very few, if any, significant conceptual or methodological gains have registered through these projects. "Some of them were poorly organised and indifferently executed.... ." The government sponsorship of social science research projects did provide an impetus to it. But on the negative side, this resulted in a great deal of mechanical and some shoddy work. The research tradition has definitely not been enriched by such studies.

Government agencies are also involved in social science research, "where one gets picture of a great deal of mechanical work and of a general sterility in respect of significant ideas, (perhaps) these organisations have some built in devices to distract them from the path of creativity." The record of some government organisations

is certainly one of high productivity, but the scale of its operations is such that quite so often, quality has to be sacrificed for quantity. Nevertheless, these organisations are a rich source of social data. A few research organisations outside the university system are also producing insightful studies, especially in regard to the developmental hazards. There is a good case for looking into and correcting the organisational incompatibilities and contradictions in the government run research institutions. Bureaucratic strangleholds and the typical frustrations of the social science intellectuals need to be examined and remedied; comparative and long-term research needs to be given special attention, organisation of research of different needs to be looked into, management guidance in this space great help.

The function of social research is mainly exploratory, cannot be expected to project a definite course of social action; it can at offer policy alternatives. That social scientist makes it possible for the policy planner to move from the known towards the desirable but the unknown is a help of major consequence. Towards this end, he has to sharpen his tools and define his focus. Needless to stress that in the final analysis the social scientists will be judged by the quality of their study-products and by their success in communicating to the community in general and to the policy planning in particular.

## **SCIENTIFIC SOCIAL SURVEYS AND RESEARCH**

It would be better to discuss what is known as 'social survey'. Until quite recently, it was customary to distinguish between social survey and social research, indeed without much justification. The contemporary trend consists, however, in viewing survey as an aspect of research and such confluent terms as 'survey-research' have found currency in methodological writings.

A social survey in its broadest sense, has a reference to a first-hand investigation, analysis and co-ordination of economic, sociological and other related aspects of a selected community or group. A survey may be undertaken with the primary purpose of formulating a programme for amelioration of the conditions of life

and work of a community or a group, implying some 'frame' in the mind of the surveyor as 'to what the conditions ideally ought to be. The purpose of a social survey may also be to provide scientifically gathered facts or materials affording some empirical basis for the social theorists to set up their conclusions.

If the term 'social survey' is mainly considered of as referring to an operation having as its central concern 'social action,' i.e., social engineering, social reform, social planning and social service, it is to the distinctive history of social survey movement that one would have to turn for explanation.

The history of the development of survey movement unfolds itself, we come across such trail-blazers as John Howard, a philanthropist and reformer; Fredrick Leplay, a reformer and economist; Charles Booth, a reformer and statistician. The life and works of all these men were governed by a deep-seated conviction that constructive reforms must be founded only on the secure ground provided by scientifically gathered facts. Thus, the survey-operation might be regarded as a pre-requisite to social reform and often the former implied the latter. Thus, the terms 'social surveyor' and 'social practitioner' came to be used almost interchangeably. It is in this particularized sense that quite a few books treat of social survey; for them it means "scientific... study of social problems acute enough to arouse public opinion and to take a hand in their solution" or a "scientific study for the purpose of presenting a constructive programme of social advance."

It is worthy of note, however, that the notion of social survey as an operation undertaken basically to afford scientifically gathered material as a basis for theory construction, found its way into the sociological thought 'as a result of Leplay's work. Threaded through the survey work of Leplay was an insistent concern, for theoretical generalization which has ever since exerted a powerful influence on the French and German sociology. Social survey with social action as its central concern while gaining substantially from the impetus provided by Leplay's work, had its actual inception with the pioneering work of Charles Booth and associates and attained its full bloom in USA. The influence of the Booth survey in the field

of social planning was as significant as that of the survey method for systematic generalizations or theory building. Thus, social survey as a method for the study and analysis of social phenomena as well as one affording a substantive base for a programme of social planning is of comparatively recent origin.

At the present time, the development of varied types of survey, both the voluntary, semi-public and the governmental agencies, marks the disappearance of social survey as a clearly defined form of social investigation having direct relevance for social planning or reform. Today, the influence of sociological surveys is not restricted to the field planning, programmes of amelioration and change. Quite a number of surveys seem to include only incidentally the programmatic.

Because of complexity of objectives and diversity of the uses of social surveys there is in evidence a lack of uniformity in defining and employing crucial variables and the qualitative and quantitative indices devised for their measurement. This has created problems, since surveys cannot be used, comparatively speaking, on a broader field of social generalization although they may be undertaken with reference to some specific problem in a specific group. It is precisely this that explains why the social survey method is no longer restricted to any particular school of thought.

It is clear from the above that a broader perspective, social survey and social research do not constitute an 'either-or' dichotomy. While a social survey may provide basis for theory construction or generalization, in addition to its implications for social planning and reform, social research may provide just the clues which may be utilized for solving certain practical problems or which may help setting up of programmes on right lines, e.g., the theory of group morale may guide planning to step up the output in a factory.

The differences between social survey and social research, as though they constituted a clear-cut dichotomy, seem to be governed by a considerably narrowed down conception of each. Looking at the matter from this angle, differentiation, even if arbitrary, is indeed easy to effect. The social survey planners are concerned with specific persons, specific places, specific problems, and situations, whereas

the social researchers are inclined to make the more general and abstract problems as their principal concern. Whereas "the social surveyor is interested in fact finding in order to improve the current social conditions of a specific locality the social researcher seeks to build a body of tested general knowledge of mankind, a body of knowledge timeless, spaceless which may lead to formulation of theories and general laws."

R.S. Lynd's differentiates, "The former (social researcher) works in a leisurely world in which the hands of the clock crawl slowly over a vast dial; to him the precise penetration of the unknown can not be hurried. In this time universe of the research scholar, certain supporting assumptions have grown up, such as... objectivity... the self-justifying goodness of 'new knowledge' about anything big or little, the practical man (surveyor-planner)... works by a small time dial over which the second-hand of immediacy and urgency hurries incessantly."

## 12

# SOCIOMETRY

There is a special technique known as Sociometry which involves collecting and organizing data relating to interpersonal relationship and sentiments.

The sociometric technique is closely identified with the work of J. L. Moreno, entitled, *Who Shall Survive?* Sociometry is concerned with charting out the attractions and repulsions among the members of a group, among groups (miniature social systems) or sub-groups or between the sub-group and individuals. Sociometry involves a set of operations that depart fundamentally from the method employed by Emory Bogardus for the measurement of social distance.

Helen Jennings, one of the pioneers in the field of sociometric studies, describes sociometry as a device for a graphic and straightforward portrayal of the total configuration of relations among the members of a group at some given point in time. Such a picture affords at a glance, the main lines of communication and the whole kaleidoscope of attractions and repulsions among members of the group. For Franz, "*Sociometry is a method used for discovery and manipulation of social configuration by measuring the attractions and repulsions among individuals in a group.*"

The basic technique employed in sociometry is the sociometric test; sociometric techniques have a reference to a cluster of devices, sociometric test included, which consist in having each member in

the group with whom he/she would like to or would not like to engage in some activity that is relevant to the life of the group. Depending on the character of the group, the members may be asked to indicate whom (from among the other members of the group) he/she would like to be associated or not like to be associated with, in play, studies, problem-solving, dinner, lending and borrowing, etc.

What type of interactions among members become the focus of the researcher's attention depends, besides his objectives, on the nature and functions of the group. Generally, sociometric studies employ observation, questionnaires and interview schedules. Sometimes, examination of records may also be employed to secure the relevant information. But sociometry should more properly be considered a method of focussing on a particular type of subject-matter and a related method of analysis rather than simply a method of data-collection.

The researcher needs to conduct observation of the behaviour of members if he wants to know the actual happenings in the group. During such an observation, the researcher concentrates on how the members behave, how they interact with one another, what the nature of their relationship is, who initiates interaction (orientation role) and who plays the object-role, etc.

Suppose, we conduct an observation aimed at ascertaining the pattern the students of a class in a public school exhibit in respect of exchanging the New Year Greetings. We may find that one student may be taken to be the most popular member of the class in as much as he receives maximum number of greetings; a sociometric 'star', to use the language of sociometry. We may also find that a few students do not receive any greeting cards. In sociometric terminology, these are the 'isolates'. We may further find that there has been among students, incidence of mutual exchange of greeting cards, e.g., A sends to B and B sends to A. This is known as the 'mutual choice'. In the course of our observations, we may come across some cliques, i.e., sub-groups of students, within each of which mutual exchange of greetings has taken place.

The sociometric questionnaires and interviews are employed in securing information from each person about the other members of the group with whom they would like to or would not like to engage in a particular type of interaction, as also, their thoughts about this interaction. Questions included in the sociometric questionnaire/schedule are directed toward seeking information from each person in the group as to which other members of the group he would like or not like as his playmate, roommate, colleague, etc. Sometimes, the person is asked to name all the persons in the group whom he would like to choose or reject; that is, there is no restriction on the respondent to confine his choices or rejections to the first three or four persons in order of preference. But if the group is numerically large, the individual respondent is usually asked to indicate his choices or rejections to the first few persons in order of preference.

*Moreno* himself insisted on soliciting unrestricted number of choices or rejections, i.e., he recommends that the respondents should be allowed to indicate the total range of choices or rejections without any limitations. There is, of course, no denying that such a freedom allowed to the individual members of the group would go a long way in affording a sensitive and objective portrait of the interpersonal relations in the group. But, practical considerations often warrant a restriction on the individual respondent to indicate his choices or rejections only upto a certain numerical limit (three or four). As has been said above, if the group subjected to sociometric analysis is a large one, such a restriction largely becomes necessary.

If, we want to administer a sociometric test in a class of students. One may ask each student of the class to indicate which three (or more) students he would like to invite for a birthday party at his place and what his order of preference would be among those he would like to invite. We may also ask each student to indicate which three (or more) students he would not like to invite to his birthday party and what would be his order in rejection.

If, you were a class teacher and wished to take the students of your class out for a picnic. Let us suppose, the hypothetical bus can accommodate only 47 students whereas the class-strength is

50. Rather than taking some arbitrary decision by yourself as to which three students should be excluded, you would naturally like the class to take the decision. One way of doing this would be to ask each student of the class to indicate which three students he/she would like to be excluded from the picnic and what his/her order of negative preference would be. This way the three students who are not liked by the students (and the information as to who is most unpopular, who comes next, etc.) may be identified.

If the respondents believe that their choices or rejections would be used as a basis for the actual restructuring of the group or for effecting subsequent arrangements or rearrangements, there is a greater likelihood of their responding to the sociometric questions in all sincerity. At least, this is the usual and reasonable assumption. So in the sociometric studies, the researcher normally prefaces his questioning with the assurance to the respondents that their choices in respect of play, living arrangements and studies, etc., Will be taken into active account while effecting modifications in prevailing conditions or in making subsequent arrangements.

Despite the fact that Moreno and associates had employed the sociometric method with extreme care and their insistence was on the use of this method only under certain conditions, the sociometric method has been used quite extensively and quite often without much caution. The popularity of this method may be attributed to the facility and graphic character of the 'sociogram' which may be aptly described as a diagrammatic means of presenting in essential basics the outline of inter-personal relations among the members of a group and the sentiments underlying these relationships. Sociometric data can also be summarized by means of  $N @ N$  table. Such a tabulation is basic to matrix analyses.

It will be better to understand the  $N @ N$  tabulation with the help of an example. Suppose we asked each one of the fifteen students who comprise a class, whom (among the other fourteen) they would like to invite for a birthday party. Suppose also, that we asked each student to indicate only the first three choices in order of his/her preference. We can present these data, in a  $N @ N$  table

(since group has fifteen students, the table will be a 15 @ 15 table).

The digits 1, 2, 3, shown in the various cells of the table below indicate the preferences. 1 represents the first preference choice. As shown in the table below, A has given first preference to B as an invitee to the birthday party. Similarly, the digits 2 and 3 in the table indicate respectively the second and third preferences. As can be seen in the table, C has given second preference to A and third preference to E. Similarly, D has given first preference to A second to C and third to E. Thus, on the whole, the sociometric matrix, illustrated above, portrays in a quite intelligible manner the interpersonal relations among the students of a class.

**Chosen**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
A		1		2		3									
B	1			2	3										
C	2	1			3										
D	1			2			3								
E	2	1	3												
F				1	3				2						
G		3			1	2									
H				2	3		1								
I	1				2			3							
J	2		1					3							
K	2		1												3
L				2			1		3						
M	3				1									2	
N						1			2	3					
O		1							3					2	
I	3	3	3	2	1	-	3	-	-	-	-	-	-	-	-
II	4	-	1	4	2	-	-	1	-	1	-	1	1	-	-
III	1	1	1	1	4	1	2	-	1	-	2	-	-	1	-
Total	8	4	5	7	7	1	5	1	1	1	2	1	1	1	0

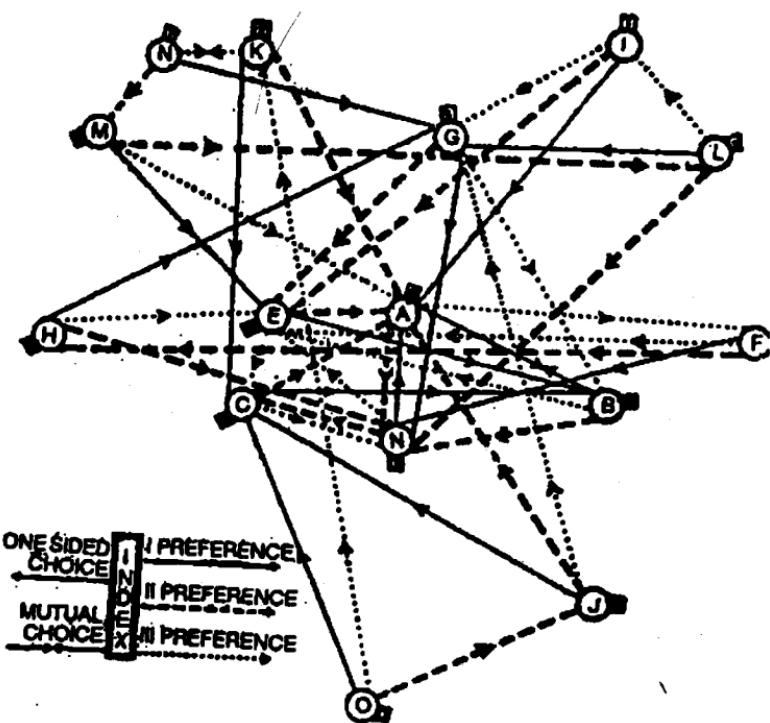
The various same data may be presented in the form of a sociogram.

## SOCIOGRAM

Examining the sociogram presented above, we find that A is the most frequently chosen person. He has secured the highest number of choices, i.e., eight. Out of these, three are first preference choices, four second preference and one third preference. In the *language of sociometry we may designate A as the 'star'* e.g., he is the most popular student in the group. In terms of the total number of choices, D and E (who have secured 7 choices each, irrespective of preferences) come next. But preferencewise, their standing is not comparable to A, since D has only 2 first preferences whereas E has only 1. It is worthy of note that C, G, and B have secured respectively a total of 5, 5 and 4 choices but each of them has received 3 first preferences. B has secured first preference choice from A who is the 'star' (as such this first preference choice, qualitatively speaking, amounts to more than any other first preference choice). Such qualitative differences are *important aspects of sociometric analysis*. In the sociogram, we find that O has not been chosen by a single student and is thus an *isolate* in sociometric parlance. Some instances of mutual choices also appear in the sociogram, e.g., A has chosen B and B has reciprocated A's sentiment. similarly, K and N have given third preference choices to each other. The sociogram also presents *mutual choices* that are qualitatively unequal, e.g., D has indicated his first preference choice for A but A has given second preference choice to D.

If there is a group considerably larger than the one represented by the sociogram, the corresponding figure might have shown the existence of sub-groups or cliques within the larger group.

The sociogram above, represents the pattern of choices with respect to invitation to a birthday party. we could have also asked the students of the class questions with a view to knowing the pattern of choices/rejections in regard to other criteria. For example, "whom would you choose as your play-mate?" or "with whom would you



like to share your lunch?" etc., are some of the questions that could be asked. It should be noted with reference to the sociometric questions that the situations or events about which questions are asked must be the ones the members of the group are familiar with and that the questions should appear realistic in the context of the group. that is, the events of situations should not appear very outlandish or far-fetched to the respondents and should meaningfully fit into the cognitive structure of respondents. Lastly, it is very desirable to maintain a fair measure of consistency between the sociometric structure we are interested in and the criteria in respect of which we ask respondents to indicate their acceptance or rejection.

The questionnaire and interviews are the principal instruments involved in the administration of sociometric tests. These are easy to administer and can be reformulated to suit situations of differing kinds. Students engaged in testing the reliability of sociometric data

have found that despite a considerable variation in the individuals' specific choices and the patterns of inter-personal relations within the group, the scores/indexes based on the sociometric data are fairly stable.

Sociometric methods have been used with advantage in the studies of leadership, friendship patterns, group structures, social adjustments minority prejudices, morale, public opinion, etc. In the field of psychiatry too (especially, group therapy) the use of sociometry has proved very fruitful.

Among the pioneering studies of leadership using the sociometric technique, the one conducted by Helen Jennings on girl students deserves special mention. Jennings calculated the choice-scores for each student on the basis of the choices or rejections received by each of them. This study revealed that there was a close correspondence between the leadership status of the students in the community and the choice-scores received by them in the study. One of the most significant findings of Jennings' study has been that leadership does not depend on any definite constellation of traits or personality characteristics; it depends rather on the behavioural contributions made by the person in reference to other members of the group.

Festinger, Schachter and Back have employed the sociometric technique in their study of the effects of residential patterning on friendship patterns. The study brought to light the fact that while factors such as age, interests, socio-economic status, etc. are important in the formation of friendly relations, ecological factors have a significant contribution to make in the initial formation and reinforcement of friendship relations. Crisswell employed the sociometric technique in her study of ethnic group prejudices. She has been able to show clearly that racial prejudices among children develop only after a certain age.

Roethlisberger and Dickson in their well-known study entitled 'Management and the Worker' have described by means of sociometric diagrams, the interpersonal relationships among workers in the

'Bankwiring room', an experimental group called out of the larger complex 'the Hawthorne Electrical Works'. Many a time, it is not feasible to study the interpersonal relations in the context of some specific concrete situation. At such times, the individuals are asked to participate in an imaginary play situation or socio-drama with a view to assessing the nature of their attitudes. The socio-drama has diagnostic as well as therapeutic implications for the participants.

**Merits:** The variables that sociometric measures represent are typically social and as such, deal with the basic data the sociologists in the main, are primarily interested in. Individuals in interaction within a miniature social system are effectively represented, affording thus, an insight into the social environment as perceived by the subject. Knowledge of the "subjective frame of reference" aids understanding of the situation. Sociologists, social psychologists and, to a lesser extent, social anthropologists and psychiatrists have employed sociometric measures with advantage. Moreno's insistence that sociometric findings should be utilized for restructuring the situation under study has led to use of these measures being employed in the fields of education, industry, military, mental health, and broadly, in social engineering.

The interdisciplinary appropriateness of the sociometric devices makes out a strong case for their relevance to a field where there is interest in cross-disciplinary integration. Furthermore, these are devices an independent researcher can use without large-scale resources. Sociometric findings have direct implications for the situation under study. This is an additional merit of these techniques in terms of action-research. Nor is it all; these techniques while generating a high degree of interest and co-operation in the subjects, also provide acceptance indices and operative definitions for a number of empirical concepts. Small wonder, the social behavioural sciences have been impressed by sociometric devices as evidenced from the frequent employment of these techniques in various areas of their scientific concern.

**Demerits :** It will be better to discuss some of the cautions and limitations a researcher employing sociometric techniques ought

to be familiar with. These cautions in some cases relate to (a) the manner of employing these techniques; in others, they relate to (b) shortcomings inherent in the very nature of sociometric techniques.

The sociometric measures provide only one of the multiple means of viewing inter-personal relations. More often than not, additional evidence gathered through other means of providing information about the physical, social, cultural and psychological aspects of the situation is a prerequisite to an interpretative understanding of the phenomena, the sociometric techniques are conventionally applied to. Moreno rightly recommends the use of spontaneity tests and interviews to yield clues to the determinants of sociometric responses. For example, a person indicating his choice for a member of his ethnic group may have done so owing to his personal acquaintance with the member chosen; not so much for reasons of ethnicity.

If, it was seen in the course of a sociometric study of children belonging to different socio-economic classes that the middle class children are generally inclined to choose children of their stratum as playmates. What would this finding mean? If we found that the (chosen and the choosers) middle class children live in the same locality and as such are acquainted with one another; this would mean that residential factor is a significant one in determining inter-personal choices. Children belonging to different socio-economic group usually live in different localities, hence the possibility of contacts between children of different strata is bound to be less. Consequently, the middle class children are understandably not inclined to choose children of other groups as their playmates. Hence, any conclusion that the act of middle class children in not choosing children of other socio-economic or cultural groups is a reflection of prejudice, would not be a reflection of reality.

This does not, however, mean that the element of prejudice as a historical fact is not involved in patterning the choices of children anywhere. For instance, in fact the segregation of people of different creeds, castes and racial stocks, i.e., their living in separate communities/localities may, in part, be attributed to prejudice.

Sociometric techniques constitute an integral part of the efforts of Moreno and associates at propounding a mature behavioural theory of inter-personal relations. The sociometric techniques acquire their distinctive character from this theory. But a few researchers may be tempted to misuse these techniques. John Madge has sounded an explicit caution bearing upon this possibility.

The tendency to ignore other kinds of data and to ascribe a special empirical status to sociometric data may retard motivation to introduce measures necessary for a fruitful and wider application of sociometry. Lindzey and Borgatta contend that there has been very little in the way of systematic, theoretically derived research aimed at exploring the kinds of conditions and variables related to sociometric response.

The relative lack of systematic and cumulative investigations is rooted in the tendency to rely solely upon the sociometric measures. Thus sociometric studies have been confined mostly to educational and institutional groups, while other pertinent groups have not been investigated sociometrically.

Multivariate conception of the nature of inter-personal relation may hopefully lead to research that will be of basic importance to the understanding of group process. Besides, such systematic research would promote greater understanding of the sociometric instruments. Lack of concern for linking of the formulations to empirical data (characteristic of Moreno) has had the effect probably of deterring a more cautious and persistent approach to relations between theory and sociometric data. Leeman has presented a base-line mathematical model to represent sociometric choices which may lead to a more systematic approach to sociometric data.

Again, the sociometric investigators have paid scant attention to the criterion of sociometric questions. The importance of selecting as appropriate criterion for sociometric analyses has been emphasized time and again. Relevant here is the distinction between 'essential' and 'auxiliary' criteria (proposed by Moreno), the essential criteria being those that refer to central activity under study. Crisswell has

distinguished between 'one-way' and 'two-way' sociometric questions while Jennings suggests a distinction between 'Psyche' and 'Socio' group. These distinctions need to be stressed since a failure to select the criterion carefully does affect the quality of data adversely and such data may not lend themselves to analysis of any consequence.

The question of what the choices under various criteria mean, as also, of the kinds of variables that can best be measured through the use of certain criteria, are questions that rightly deserve careful consideration and yet have often been neglected. One of the objections raised in this connection is that the investigator employing sociometry does not often know what a choice means for the subject. This objection refers to the question of relevance of various criteria for providing data in regard to particular dimensions.

The demand that the investigator should "know" what a response actually means to every subject is not very proper; the important thing, however, is the extent to which an order imposed upon subjects on the basis of response can be related to pertinent independent measures.

Despite the phenomenal advances made in the analysis of sociometric data in the last few decades, characteristic flaws still persist, the most serious of these being the "tendency to capitalise illegitimately upon chance-variation" and to treat such variations as socially significant.

'Near sociometric' and 'quasi-sociometric' studies which are ramifications of the 'pure' sociometric studies, have also been conducted quite fruitfully in situations where strict conformity to the requirements of 'pure' sociometric studies is beset with practical problems. Sociometric self-rating, group participation scale, multirelational sociometric survey and 'Guess who technique' etc. are some of the methodological devices, are similar to the sociometric tests.

## **RESEARCH IN LIBRARY AND INFORMATION SCIENCE**

With the constant efforts of Dr. Ranganathan, the discipline of librarianship took the form of library science, as it adhered all the norms and principles of scientific research.

In the last phase of 20th century, many Indian Universities having library and information science department which were imparting the courses of Master of Library Science, on the insistence of Dr. Rnganathan started doctoral course in the area of library and Information science leading to the degree of Ph.D. Very few cases are available to the degrees leading to D. Litt. in library science as well.

Recently, IGNOU has announced a programme for working professionals to go for doctorate through Distance mode of education.

The brief description of various methods in the library and information science are discussed in the following paragraphs.

The research setting and design is a conceptual system within which research is conducted. It constitutes the blueprint for the collection, measurement and analysis of the data. As such it includes an outline of what the researcher will do from writing the hypothesis to its operational implication in the final analysis of data. Prior to research following decisions are to be taken :

- (a) What is the study about ? (Identification of problem).
- (b) What kind of method to be employed by means of which the identified problem will be studied ? (Methodology).
- (c) Why is the study being made ? (Purpose of study).
- (d) What type of data is required ?
- (e) From where can required ?
- (f) What period of time will the study require ?
- (g) What will be sample design ?
- (h) What Technique of data collection to be used ?
- (i) How to employ a particular method to carry out research successfully ?
- (h) In what style will the report be prepared ?

Research setting and design is needed to facilitate the smooth setting of various research operations, to make the research as efficient as possible, yielding maximum information with minimal expenditure of efforts, time and money.

The research methods applicable in library and information science have been classified as under :

#### **Research Methods in Library and Information Science**

<i>Historical Method</i>	<i>Survey Method</i>	<i>Case Study Method</i>	<i>Delphi Method</i>	<i>Statistical Method</i>
Observation Technique	Questionnaire Technique	Interview Technique	Documentary Technique	

#### **Historical Method**

According to the Shorter Oxford English Dictionary, history is "a written narrative constituting a continuous methodical record, in order of time, of importance or public events, especially those connected with a particular country, people, individual etc. "History is essential, if we are to understand our present environment, then we should know how it grew up. History is the story of its development;

of its evolution and of its origin, and it helps to explain the present in this way.

History of library and information science is the continuous methodical recounting of past events pertaining to the establishment, maintenance and utilization of systematically arranged collection of recorded information. It narrates, in order of time, the various developments, their causes and effects on the professional of librarianship.

Goldhor considers historical research "a battery of devices by which one is added to reach a conclusion as to the probable truth of events in past, as reflected in the objects available for study in the present". Thus historical research method is a past oriented research which seeks to illuminate a question of current interest by an intensive search of materials, whose existence had not previously been known to recent generation. The discoveries made by various scientists by re-interpretation of events in the light of the increasing amount of information available, is also historical research. Thus truth is that a proper study of past will provide key to the present. The present day systems and its varying forms and rules could be easily explained by the study of the history. Although the present is all together different from what the past was, but still the origin of the present lies in the past.

Historical method, like in several other disciplines, if adopted in Library and Information Science research, will be foundation for improvement on the basis of the past study. It can also help to find out various shortcomings in the field on the basis of past. It "can contribute to the body knowledge about librarianship and it can also facilitate our understanding of when, how and why past events disseminations". It also often recognises the social economic, political, intellectual and cultural environment in which these events occurred. This method can be used to study (a) Historical development of libraries ad library materials, (b) Impact of printing on libraries, (c) Development of library profession, (d) Development of education for librarianship, (e) History of library systems and services, (f) Role of library associations, (g) History of library movement, and

(h) Emergence of multidimensional subjects and their impact on library profession.

## SURVEY METHOD

A survey is "a systematic collection of data concerning a system, its activities, operations, persons involved in that system, also persons who are involved in that system." The man "purpose of the survey is to collect, organise and all the surveys which almost inevitably calls for providing evidence that the deficiencies exist."

The library survey may be defined as a systematic collection of data concerning a library, its activities, operations, personnel working in the library and its users. The purpose of library survey is to make a specialised type of investigation to improve library services. Library surveys are conducted either to assess an existing situation or to check library system or to evaluate the area of librarianship in order to remove shortcomings or inadequacies and to plan further.

The vital aspect of survey method is collection of data. Following techniques are mainly employed for data collection in survey method of research :

### **Survey Method**

1. Observation Technique
2. Questionnaire Technique
3. Interview Technique
4. Documentary Technique

### **Observation Technique**

Observation is at once the most primitive and the most refined of modern research techniques. It is, undoubtedly, the first procedure of science. It underlies all research; it plays particularly a prominent part in the survey procedures. P.V. Young defines observation as "Systematic viewing, coupled with consideration of the seen phenomena, in which main consideration must be given to the larger unit of activity, by which the specific observed phenomena occurred". Observation is a perfect method of serial investigation. For field

data the observation technique is also called classical technique. It is a process of gathering information by means of direct sensory perception. A relationship is established between researchers and respondent, in this technique. It enables a direct contact between subject and object involving intellectual processing of the subject or any part of it with a definite purpose. This technique can be used to study (a) users needs, (b) user behaviours, (c) users reading habit, (d) use of library material, (e) evaluation of library services, (f) functioning of different sub-systems of a library system, and (g) users approach to information.

The observation may be 'Participant' or 'non-participant'. When the researcher participates in the activities of the group under study, it is known as participant observation. But when the researcher abstains from the activities of the group, and simply observes them from a distance, it is known as non-participant observation. Purely non-participant observation is extremely difficult. Again the observation may be 'controlled' or 'uncontrolled'. When the observation is made in the natural setting, and the group activities are carried out in their usual course, without being influenced or guided by any external force it is known as uncontrolled observation. In controlled observation group activities are guided and phenomena is put to laboratory type test, under guided conditions.

### **Questionnaire Technique**

In fact I should have used the term 'mailed questionnaire' since the instructions denote that the term questionnaire is used as a guide in interviewing. Questionnaire is "a formal list of questions, especially as used in an official enquiry". In this technique, to collect data essential for conducting research, a set of questions are asked from the participants. Questionnaire is constructed translating the aims and objectives of the survey study. This is a major instrument for data gathering in survey studies. The literature survey reveals that most of the research is conducted by this techniques in its more complete coverage of users' interests and analysis of their replies, which appear to be relatively simple procedure.:

This technique is most feasible and economical for gathering data when the coverage of study is geographically very large and where researcher could not collect it personally. This technique can be helpful for research studies like (a) users needs, (b) evaluation of library services, (c) users survey, (d) nature and extent of library resources, (e) employment prospects, salaries, job satisfaction of library staff, (f) functioning of library system, (g) study of library automation, and (h) study of cooperation among libraries etc.

The main advantage of a questionnaire technique is low cost for a large coverage. It is suitable for repetitive information with greater validity.

### **Interview Technique**

A research technique very similar in nature and purpose to the questionnaire is the interview. In fact, except for certain relative advantages which need to be clearly recognised the two techniques are, for some purposes atleast, essentially interchangeable.

Interview technique is a direct method of data collection. It is a conversation carried out with the definite purpose of obtaining certain information by means of the spoken words. It may be regarded as "a systematic method by which a person enters more or less imaginatively into the life of a comparative stranger". Thus its importance arises from the necessity to come into contact with individuals to get access to facts and opinions.

This technique is suitable for the following areas of research (a) user satisfaction survey, (b) users needs, (c) library staff job satisfaction, (d) employment prospects, status in librarianship, (e) library co-operation, (f) library resource-sharing, (g) Inter-library loans, (h) manpower planning, (i) union catalogue, and (j) library administration etc.

The interview can be either 'structured' or 'unstructured'. In structured interview complete schedule is used. The interviewer has to abide by the questions of schedule. Whereas in unstructured interview no direct or predetermined questions are used. The field worker may be told certain broad topics upon which the information is to be collected.

## **Documentary Technique**

Besides observation, questionnaire and interview, there is another source of data which may be termed as documentary source. It has been an important source of information. Through documentary technique the researcher makes use of any or many documents or record, published or unpublished to extract necessary information. Document is a very important, dependable and valuable source of information which is a record that contains important information about a problem or aspects of study. Document may be a life history, diaries, letters, official ad unofficial records, proceedings of parliaments, committees, courts, societies etc. This technique can particularly be used for data collection in following areas of research (a) Bibliometric studies, (b) Historical studies, (c) Indexing and abstracting, (d) Biographies, (e) Thesauri construction etc.

Literature available on the topic and observation reveal that none of the above four survey techniques of data collection-observation, questionnaire, interview, and documentary-is complete in itself. For greater validity of research results, these technique should be used in combination.

## **CASE STUDY METHOD**

Case study is "a technique in which an individual or group, institution or phenomenon is recognised as a unit of study and various aspects of the unit are studied deeply". Unit of study may be an individual, a family, an institution, a culture group or an entire community. It may also be an abstract thing like a set of relationships or processes viz. administrative crisis in libraries, system adjustment, library cooperation etc. It aims at deep and detailed study of the unit, though the field of study is comparatively limited. It aims at studying everything about something, rather than something about everything as in the case of statistical methods.

In essence the case study resembles almost all other types of research, for instance, in the sense that the present case can be understood only in view of its past. It is closely related to documentary research in that it deals with living individuals in their present set

up. Case studies resemble survey studies in that they are concerned with the present status of phenomena. They differ from survey studies, however, in that the determination of status is only a secondary aspect in the situation; the more fundamental question is discovering how it got that way.

In case study methods emphasis is on principles and processes rather than the transfer of factual information. It presents real solutions drawn from practice and provides an opportunity to enquire skills in analysing problems, making decisions and solving them. It helps in developing technical skills essential for librarianship. This method can be used in (a) library administration/management, (b) system analysis studies, (c) cost benefit analysis studies, (d) cost effectiveness analysis studies, (e) library effectiveness analysis studies, and (f) computer application in libraries.

## **DELPHI METHOD**

Delphi method is basically a technique of obtaining consensus among experts opinion on a given problem. A questionnaire is prepared translating the aims and objectives of research. The identified problem is put up to the panel of experts in many rounds till a consensus agreement is achieved. According to Halmer "Delphi method is a technique for obtaining opinions from a panel of experts in a particular field, through a questionnaire and these questionnaires are submitted several times until a judgemental data is obtained." The basic theory behind this method is that consensus opinion among majority of opinions will have greater credibility and authority than the surmise of only the most articulate / spokes-persons in a group of participating respondents. "The Delphi study is performed by sending a systematically formulated questionnaire to a pannel of experts in a number of rounds."

The Delphi study is performed in following rounds :

*First Round* : An open ended questionnaire is prepared and sent to the selected experts in panel.

*Second Round* : The panel members are asked to rank order preliminary priorities among the responses.

*Third Round* : The panel members are now asked to vote on the ranks of the items and asked to provide a statement of the reason.

*Fourth Round* : Final answers are requested on all the arguments and counter arguments that were presented.

The medians of the responses of this final round are accepted as the group's opinion, representing the nearest thing to a consensus of the problem. The following areas of research can be studied through this method (a) Formulation of library legislation, (b) Library policies making, (c) curriculum design for librarianship, (d) Methods of teaching and evaluation, (e) Decision making process, and (f) Manpower planning and requirements.

## STATISTICAL METHOD

Statistics is the science which deals with the collection, classification and tabulation of numerical facts as the basis for explanation, description and comparison of phenomena. And "Statistical method is basically a technique of handing quantitative information in such a way as to make that information meaningful. Statistical method is concerned with numerical relations of those aspects of things which repeat themselves indefinitely."

Statistical methods are being used for aggregative analysis and intensive study of individual unit is outside its scope. Statistical inferences are based upon averages and they are applicable upon the population as a whole and not on any particular as unit. This method is based upon quantitative analysis. Only the phenomena which is capable of being expressed numerically can be studied numerically. Thus in the process of conducting research researchers have to compile and collect different types of numerical data. Unless techniques of quantitative analysis are used, the data, too, will not be useful.

By using statistical methodology, the researcher can test hypothesis, compute means and other measures of central tendency; assess the relationship between one variable and another; make predictions determine the reliability and validity of instruments and

measurements; and generalise conclusions from sample data.

This method can specifically be used to study the following problems : (a) Budget estimates of library (b) Library planning studies, (c) Assessment of library services, (d) Evaluation of library services, and (e) Library forecasting studies etc.

## SCIENTOMETRY/BIBLIOMETRY

The bibliometric studies in India started with the publication of an article by S. Dutta and T. S. Rajagopalan in 1958. However, it may be noted here that the term librametry was coined by Ranganathan in 1948 during a discussion in Aslib Conference held in Lamington Spa (Aslib Proc. 1949, 1, 102), and the term was used more or less in the same sense of bibliometrics in India for quite sometime.

Bibliometric studies in India took a firm root in 1963 when a seminar was organised by DRTC on Documentation Periodicals and quite a few papers were presented in the conference including foreign ones. Eversince, the studies have been continuing. During the period of the last 27 years, about 200 papers have been published on the coverage and overlapping of literature, ranking of periodicals, use of literature by scientists and others, testing of various bibliometrical laws and so on. A survey of the studies conducted is long overdue to find out what has already been done, how they have done, what remains to be done, whether there is some gap in our studies etc.

The year 1988 is recorded in the history of bibliometric studies in India as one of the memorable years in as much as this particular year witnessed the establishment of the Bibliometrics Section, and the National Centre on Bibliometrics, at INSDOC, organization of a 12-day work-shop on bibliometrics and scientific communication

IICB, Calcutta, and a seminar by the Society of Information Science, also at Calcutta on the same topic. In addition some noteworthy studies have been conducted and a sizeable number of good papers published. Three decades that have passed eversince the publication of first paper on bibliometrics from India, no year has been so eventful as the year under was 1988.

Visualising the potentialities of bibliometrics as the effective tool for analysis of various research outputs, and generation of indicators that may prove highly useful for planning, decision making, and various other functions of science management, Dr. A.P. Mitra, Director General, CSIR entrusted INSDOC in May 1987 with the task of carrying out bibliometric analysis of CSIR research output of 1986. Mr. T.K. Dutta, the then Scientist-in-Charge of INSDOC constituted an informal Bibliometric Group with the primary author of the article as head and provided the necessary impetus to carry out the job. The job completed in a little over 3 months, generated a number of useful indicators and earned unreserved praise of the DG, CSIR and evoked great interest in CSIR. The Group was then, assigned the task of carrying out the bibliometric analysis of selected papers, numbering more than 3,000 from CSIR laboratories. The objective was to identify the significant papers published by CSIR scientists since independence, which were to be included in the Commemoration Volume projected to be published in the 40th year of India's independence. Alongside the aforesaid job, the bibliometric analysis of research output 1981-85 was also taken up. At this stage it was realised that a lot of bibliometric studies need to be carried out, and the activity has to be ongoing. Hence the Bibliometrics Section was formally created in February 1988, with the same group informally constituted before.

## NATIONAL CENTRE ON BIBLIO METRICS

While accomplishing the aforesaid jobs, some of the problem of bibliometric analysis of Indian Science and Technology research output became evident. The coverage by Science Citation Index (SCI) of Indian periodicals being very small (18 in 1986), the

determination of impact factors of large number of India S and T periodicals not covered by SCI, created lot of problems. Method had to be evolved for the same which has since been perfected.

The citation analysis of research papers published during the last 40 years unveiled the inadequacy of bibliometric tools available in Delhi. Not a single library in Delhi was having the complete run of SCI from 1955 to date. The back runs of 3 libraries in Delhi situated far away from one another formed the complete set since 1966. The 1955-64 culmination of SCI, was not available in any of the Delhi libraries.

Though a sizable number of bibliometric studies are going on in India, there is no central agency to monitor and, if possible, to co-ordinate these studies. Training facilities for bibliometrists are also scarce.

All these factors led to the establishment of a national Centre on Bibliometrics, with the support of NISSAT, and it is functioning since April 1988. The objectives of the Centre are :

1. To create S and T citation data base of Indian contributions appearing in Indian journals. The comprehensive collection of research journals in the National Science Library will be the base for creation of such data base the format could be the same as that developed for Science Citation Index.
2. To develop tools, techniques, and modalities for the analysis of research output based on SCI data and Indian S and T citation database to be created.
3. To analyse the research outputs of selected research institutes, agencies, universities and other similar cognate bodies.
4. To conduct short training courses in bibliometrics and its applications.
5. To render bibliometric services to the scientists and technologists of the country, etc.

Basu and Garg (1999) conducted a study on changing trends in Indian Bibliometrics. The following paragraphs are based on this study.

## EARLY PERIOD OF BIBLIOMETRIC STUDIES

Since the 1970's the term bibliometrics has been increasingly used to describe mathematical and statistical analysis of literature. The area has the characteristics of both basic and applied research. For example, the quantitative relationships that were noticed in the early stages followed from empirical work and related to patterns in the productivity of journals, productivity of individual scientists, or word frequencies in the literature. These relationships that go by the name of Bradford, Lotka and Zipf, all collectively come under what is now generically referred to as the '80-20 Rule'. In other words, what was observed was that productivity patterns are usually highly skewed such that approximately 80% of the published papers in a field are located in 20% of the relevant journal set, or, 80% papers are authored by 20% of all authors. A similar pattern exists in word use in textual matter, where a power-law relationship is observed between the frequency of a word and its 'rank'. These patterns have been theoretically analyzed in terms of mathematical models to identify and explain the complex underlying processes leading to the specific behaviour. Such aspects are now sometimes referred to as 'Informetrics' since they are supposedly 'information-related'. The empirical relationship of Bradford law found of immediate use in the economics of library acquisitions. Similar relations were found to be useful in other aspects of library management such as shelving, access, issuing of books, etc. since the first practitioners of bibliometrics were those in the profession of Library Science. This initial historical bias remains to this day to some extent.

In the early period, bibliometric studies were sporadic and often conducted by individuals who wished to project the state of a particular discipline or group of performers. At this time, a cohesive group of practitioners with measurements identity and their own journals and discussion forums had not emerged. With the establishment of the journal *Scientometrics* from Hungary in 1978, and the initiation of a regular biennial international conference since 1987, the situation has been somewhat rectified, and a cohesive research community may be said to have around the fledgling discipline.

## **EMERGING TRENDS IN BIBLIOMETRIC STUDIES**

Within the last decades, the nature and scope of bibliometric studies has evolved considerably, moving away from its initial foundations in library science and becoming more policy orientated. A definitive direction had been given to the discipline by the seminal work of de Solla Price in 1963, which may be considered as the precursor of what are called scientometric studies today. The possibility offered by the record of publications to formulate indicators that could give measures of performance and activity in the complex system of knowledge production and exchange that constitutes scientific activity was first reviewed in 1978. In these quantitative studies of science, sophisticated data collection and data-handling techniques played a significant role. The possibility of developing 'maps' emerged, using techniques such as co-citation or co-word analysis which exploit associations to elicit the similarities or differences between disciplines, research groups, country profiles, etc. Networks of authors and papers could be traced, in addition to tracking of emergent disciplines. The field developed under dual pressure; (i) from the science management system to evolve objective quantitative measures that could be used in policy formulation, and (ii) by the creation of new possibilities offered by computerized data and computer based analytical tools. Certain fundamental issues in the use of indicators, ethical aspects such as the of publication based indicators for individual assessment, problem sessing science in general, and the use of bibliometric indicators in developing countries, in particular, have also been addressed. Methods and techniques have been developed to study the cognitive processing development of scientific disciplines, and used in historical and studies, for example, age-productivity relationships. The study also focused on technology by analysing patents rather than convent scientific literature. As the issues moved away from those specifically suggested for this area of studies. These terms along with bibliometrics are used at present in a equivalent and interchangeable sense implying statistical analysis of bibliographic data leading to inferences regarding scientific and technological activity.

## BIBLIOMETRICS IN INDIA

After the early tradition of quantitative studies in Library Science stated by Ranganathan, increased activity followed a seminar organization 1963 by the Documentation Research and Training Centre (DRTC) the Indian Statistical Institute in Bangalore. During the last two and half decades, there has been a growing awareness about this field, as witnessed by the introduction of courses in bibliometrics in schools library Science, publication of papers, organization of conferences and the establishment of the National Centre for Bibliometrics at the Indian National Scientific Documentation Centre (INSDOC) under the Council of Scientific and Industrial Research in New Delhi in 1988.

The publication trends indicate active participation by a group workers in scientometric and bibliometric research. The number of annual publications in our country has grown from 31 in 1970-74 to 154 in 1990-94 indicating a doubling time of less than 10 years. This is comparable to the growth rate of publications in this area in the world (Table 14.1). Earlier reviews on bibliometric/scientometric studies in general and with reference to India in particular may be seen in Sengupta Sen and Kumar, Sen and Chatterjee, Ravichandra Rao and Kabir.

The field of scientometricians now holds a biennial international conference where India participates on a regular basis. In the first two conferences held in Belgium and Canada the Indian participation was limited about two participants. India hosted the third conference at Bangalore 1991 in which there were 11 Indian contributions. At the fourth conference at Berlin, there were 23 Indian delegates out of a total of 138 participants from 32 countries. This was the highest participation from any country barring the USA and the host country Germany. In the fifth and sixth conferences held in Chicago in 1995 and Israel in 1997, Indian participation was limited to 11 and 5 delegates respectively.

**Table 14.1: World Output Vs Indian Output in  
Bibliometrics/Scientometrics**

Year	World output	Indian output journal	Indian output conferences	Indian output other	Indian output total	% of world output
1970-74	147	28	2	1	31	21.08
1975-79	287	26	1	1	28	9.75
1980-84	412	81	26	3	110	26.69
1985-89	520	91	40	3	134	25.76
1990-94	603	104	50	-	154	25.54
Total	1969	330	119	8	457	23.20

### **TRENDS IN SCIENTOMETRIC RESEARCH IN INDIA: 1970-1994**

Bibliometric studies performed in the first two decades related to library and information science. Many of these were case studies that have appeared in Indian journals of library and information science. Emphasis has been on library related issues such as the scattering of literature of specific disciplines in journals (or ranking of journals) and the extent of citation of journals in reference lists of articles. Since 1990, the situation does appear to be changing, with a number of theoretical studies the emerging use of bibliometrics for decision making, and national and policy issues being reflected more often. The application of new mathematical and statistical techniques and the use of computerized databases has been initiated but needs to be further strengthened. There appears to be a need for a coordinated program on scientometrics/bibliometrics studies to be performed in India in order to be of use of decision making at the national level.

Before undertaking a content analysis to substantiate these statements, we examined the overall characteristics of scientometric output from India taken from a 20 year bibliography on Indian contribution scientometrics/bibliometrics by Gupta (1970-90), supplemented with bibliographic data for the period (1991-94) from *Library and Information Science Abstracts (LISA)* Conference

papers were included, but not papers prior to 1970 as their number was small and their relevance to the present study fairly limited. Some additional journals such as *Journal of Scientific & Industrial Research* (JSIR), were included as several scientometric studies have been published in these journals.

The number of papers published from India and the world, during the period 1970-94, listed in Table 14.1, shows that the national output has grown considerably, and the percentage contribution of Indian output to the world output may be higher than in other disciplines. Within a period of 20 years, Indian output has increased from 31 in 1970-74 to 14 in 1990-94. The ratio to world output is fluctuating but appears to have remained roughly at about 25 per cent since 1984. Of the 457 papers including in the study, 330 were published in journals and 119 papers were presented at conferences close to half of these being presented at the five international conferences on informatics since 1989, and the rest in conferences held in India (Table 14.2). The remaining 8 items were book chapters.

When the publications were grouped into scientific disciplines, it was found that the maximum number of scientometric/bibliometric studies was undertaken in the field of Biological Sciences followed by Earth and Space Sciences, Agricultural Sciences and Food Sciences. Only a few studies have been undertaken in the areas of Environmental Sciences, Energy or Engineering and Technology. Physical Sciences, Chemical Sciences and Medical Sciences were all equally represented. By and large, sub-disciplines received more attention than the entire discipline, except in chemistry which received as much attention as the micro disciplines.

In terms of bibliometric aspects, e.g.

- (i) Country-wise analysis of output
- (ii) Discipline-wise analysis
- (iii) Scattering of articles in journals or journal ranking
- (iv) Choice of journals for publication
- (v) Citation of journals
- (vi) Author productivity

**Table 14.2. Indian Contributions in National/International Conferences**

<i>Conference</i>	<i>Lecture</i>	<i>Date</i>
International		
Fifth International Conference on Scientometrists, Bibliometrics and Informetrics,	Chicago, USA	1995
Fourth International Conference on Scientometrists, Bibliometrics and Informations	Berlin, Germany	1993
Third International Conference on Scientometrists, Bibliometrics and Informetrics	Bangalore, India	1991
Second International Conference on Scientometrists, Bibliometrics and Informetrics	Ontario, Canada	1989
First International Conference on Scientometrists, Bibliometrics and Informetrics	Brussels, Belgium	1987
National		
Proceedings of Seventh SIS Annual Convention	Calcutta, India	1988
Science & Technology Indicators for Development	New Delhi, India	1987
Bibliometric Studies and Current Information	Calcutta, India	1985
Primary Communication in Science and Technology	New Delhi, India	1981
Other 15 Conferences		
Total		119

- (vii) Collaboration patterns, and
- (viii) Growth of scientific publications

It is found that there is a predominance of scattering or ranking studies and citation analysis, mostly published in journals devoted to library and information science reflecting the bias of library schools where a majority of these studies were conducted. About two thirds of these studies appearing in Indian journals and the remaining in international journals (Table 14.3).

**Table 14.3. Distribution of Papers in Indian and Foreign Journals**

Year	Papers in Indian journals	Papers in foreign journals	Total	Average annual publications	% of papers in Indian journals
1970-74	17	11	28	5.6 (3.4)	60.7
1975-79	20	6	26	5.2 (4.0)	76.9
1980-84	67	14	81	16.2 (13.4)	82.7
1985-89	61	30	91	19.2 (12.2)	67.0
1990-94	76	28	104	20.8 (15.2)	73.1%
Total	241	89	330	13.2 (9.4)	73.0

Figures in parentheses indicate publications in Indian journals.

The most active authors in the field are listed in Table 14.4 where we note an increasing trend in author productivity. The average annual author productivity was seen to increase from 0.4 in the first decade 1970-79 to 0.6 during 1990-94. The top 10 authors produced about 45 percent of the total output (Table 14.4). In Table 14.5, the distribution of papers by these authors in Indian and international journals is compared to that for less prolific authors. It was observed that the latter group published less in international journals.

**Table 14.4. Journal Publications of Indian Authors in Scientometrics/Bibliometrics**

(Total 4 papers or more; only first authors have been considered)

<i>Authors (No. of papers in Indian journals)</i>	<i>1970-79</i>	<i>1980-89</i>	<i>1990-94</i>	<i>Total</i>
Sengupta, I.N. (17)	12	17	3	32
Gupta, D.K. (21)	1	17	6	24
Arunachalam, S. (5)	1	12	2	15
Gupta, B.M. (12)	8	4	1	13
Maheshwarappa, B.S. (12)	-	9	3	12
Garg, K.C. (5)	-	4	6	10
Subramanyan, K. (0)	5	4	-	9
Sen, S.K. (3)	-	4	4	8
Munshi, Usha (5)	-	-	5	5
Sen, B.K. (2)	-	1	4	5
Nagpaul, P.S. (0)	-	-	4	4
Kalyane, V.L. (4)	-	-	4	4
Mahapatra, G. (4)	-	-	4	4
Sinha, S.C. (4)	-	1	3	4
Narayan, S (1)	2	1	1	4
Rao, I.K.R. (2)	2	1	1	4
Rangarajan, K.S. (3)	1	3	-	4
Nag, D.K. (4)	-	4	-	4
Raina, Roshan (4)	-	4	-	4
Average annual author productivity of more productive authors	0.40	0.57	0.72	0.36

**Table 14.5 Publication of All Authors Vs Prolific Authors in Indian and Foreign Journals**

	<i>Papers in Indian journals</i>	<i>Papers in foreign journals</i>	<i>Total</i>	<i>% papers in Indian journals</i>
Prolific authors	96	57	153	62.7
Others	116	27	143	81.1
Total	212	84	296	71.6

## CONTENT ANALYSIS

For changing trends, a content analysis of the titles and abstracts of the 450 plus individual article published from India over twenty five years was done. The papers were group in terms of their general theme into the following categories:

- (i) Library issues
- (ii) Theory/modelling
- (iii) National level studies
- (iv) Policy related studies
- (v) Case studies
- (vi) General studies

The last category included classical bibliometrics and the 'laws'. The number of papers relating to each of the above themes are noted in 5-year time blocks. The distribution of papers in each theme during different time intervals is shown in Table 14.6.

It is seen that while the number of papers in each 5 year block have been increasing overall, papers related to library issues grew in the initial years but have been declining since 1984. On the other hand, the number of theoretical studies have been steadily increasing. During the decade 70's macro level studies on national issues were hardly taken up. Since the 80's national issues have been more frequently addressed along with Theory/Modelling. Policy studies were absent in the first decade and have just emerged. The total output from India has been dominated by a large number of case studies, though these appear to be declining since 1990 after an initial stage of proliferation.

This indicates that there has been a qualitative change with more studies being addressed to national or theoretical issues rather than isolated case studies. The emphasis shifted from library related issues other areas related to performance evaluation and policy. The initial proliferation of illustrative case studies, together with an increasing number of studies related to library issues and national level studies may be linked with the higher percentage of paper in

**Table 14.6. Distribution of Papers in Terms of Content and Type of Study**

	<i>1970-74</i>	<i>1975-79</i>	<i>1980-84</i>	<i>1985-89</i>	<i>1990-94</i>	<i>Total</i>
1. Library Issues	6	8	23	12	10	59
2. Policy/Evaluation	-	-	4	10	7	21
3. Theory/Modelling	4	2	13	16	25	60
4. National Issues	4	4	30	24	38	100
5. Case Studies	13	8	32	65	34	152
6. General Studies	4	6	8	7	6	31
<b>Total</b>	<b>31</b>	<b>28</b>	<b>110</b>	<b>134</b>	<b>120</b>	<b>423</b>

Indian journals, which was as high as 80 per cent. Some of the case studies published in the national journals and using limited data e.g. pertaining to a single institution or a single journal would have resulted from student exercises or dissertation rather than more professional research from student reports. This is also partially corroborated by the pattern of publication by the more productive authors, who may be identified as the continuing professionals in the field, who publish significantly more in international journals (Table 14.5).

In India, the number of studies relating to policy matters is still fairly small, but may be expected to increase if there were demands for scientometrics inputs from decision making bodies. These changes are in line with international trends where bibliometrics/scientometrics is being increasingly seen as a tool for formulating national productivity indicators and for decision making rather than as a part of library science.

Other inferences from our examination of the themes are that sociological studies relating to relating to research environment, productivity, etc. using bibliometric inputs are still few in number. Studies using large scale computerized databases of publication output, and computer based statistical analysis and sophisticated multiversities concepts had just been initiated around 1994. Although isolated national issues have been addressed, a comprehensive mapping of national level scientific output on a regular basis is awaited.

## CURRENT DIRECTIONS IN SCIENTOMETRIC RESEARCH IN INDIA AND OTHER ISSUES

This study indicated that in the early years scientometric/bibliometries in India during the period 1970-1990 have generally been context of or within the realm of library and information science, the largest number of studies on scattering of literature in journey ranking and extent of citation of journals. Much of the work done in field represented individual initiative, with the result that there was perceptible thrust or direction. Often the data set used for the stay were meager and analytical tools used so nominal that those could be considered as illustrative exercises rather than representative

case dies, and the results could not be used for any real-life decision making Comprehensive studies at the national level were few. In other, there was a need to make a transition from "little scientometricians" to scientometricians". From the 80's, the nature of the studies has been changing with fewer studies being addressed to library issues and theoretical studies, and national and policy questions being reflected more. This shows a growing awareness in the research community about the potential of scientometric studies, in line with world trend.

Many countries such as Mexico, Hungary, Spain, Australia have expended considerable effort in mapping scientific profiles of their respective countries, along with assessment of performance. Within the last year, a need was felt to begin a similar coordinated program for India. It was suggested that if we are to enter into scientometric assessment of Indian science in a meaningful way, then one must create and sustain an information base from which overall trends in-science in the country could be assessed and monitored. With this understanding, a national level study project, National Mapping of Science in India was sponsored by the Department of Scientific and Industrial Research and the National Information Systems on Science and Technology (NISSAT). The first set of reports on this project were completed in 1998 and were based on SCI and other subject specific databases including Indian Science Abstracts.

Thus, the trend toward the use of comprehensive computerized databases has been initiated, although their use is not yet widespread. At the same time, efforts are being made to induct new techniques and statistical methodologies into this studies. The area that needs emphasis is technology tracking using bibliometric techniques.

A second point relates to the visibility of the scientometric studies to a wider community of scientists. Most of the articles published during 1970-90 appeared in journals related to library and information science (Table 14.7). The restricted nature of the readership of library science journals would have certainly influenced the visibility of these studies in a wider scientific community. As a

**Table 11.7. Choice of Journal for Indian Contributions to  
Scientometrics 1970-1994**

<i>Journal</i>	<i>Country of publication</i>	<i>Impact factor</i>	<i>Paper</i>
1. Annals of Library Science and Documentation	India	Nil	71
2. IASLIC Bulletin	India	Nil	41
3. Scientometrics	Netherlands	0.593	35
4. Library Science with a Slant to Documentation	India	Nil	18
5. I.L.A. Bulletin	India	Nil	14
6. Journal of Scientific & Industrial Research	India	0.237	13
7. Journal of Library and Information Science	India	Nil	12
8. Library Herald	India	Nil	12
9. International Information, Communication & Education	India	Nil	10
10. Herald of Library Science	India	Nil	9
11. Journal of Information Science	UK	0.224	8
12. International Library Review	UK	0.224	7
13. Other 5 Indian Journals (with < 5 papers)	India	0.205	<5
Current Science			
Indian Journal of Agricultural Lib. & Information Sci.	India	Nil	<5
Lucknow Librarian	India	Nil	<5

Science Age		India	India	<5
Science Today		Nil	Nil	<5
14.	Other 14 foreign journals (with <5 papers)	34	34	<5
Bulletin of IAALD	USA	Nil	0.330	<5
Czech Journal of Physics	Czech	2.603	<5	<5
Environmental Science and Technology	USA	0.386	<5	<5
IEEE Transaction on Engineering Management	USA	Nil	<5	<5
IEEE Transaction on Professional Communication	USA	0.670	<5	<5
Information Processing and Management	USA	0.670	<5	<5
International Forum for Information				
& Documentation	Netherlands	0.246	<5	<5
Journal of American Society of Information Science	USA	1.074	<5	<5
Journal of Documentation	UK	1.033	<5	<5
Nature	UK	25.466	<5	<5
R & D Management	UK	0.043	<5	<5
Special Libraries	UK	0.535	<5	<5
Unesco Bulletin for Libraries	USA	Nil	<5	<5
World Patent Information	UK	Nil	<5	<5
Book Chapters				
15.	Conference Papers (Details given in Table 2.3)	8	8	119
16.				119
Grand Total				423

result, the role of scientometric studies for obtaining an overview of a discipline, and as an index of health, productivity and impact of a research area has not been brought out until quite recently in the Indian context.

Finally, there are several issues with regard to using bibliometric performance indicators that need to be debated jointly by scientists along with those who generate these indicators. These relate to the meaningfulness or otherwise of the indicators, their limitations, regions of validity and conditions under which they may be deemed acceptable. Wholesale rejection or uncritical acceptance of scientometric methods of evaluation, without regard to the methodological or ethical issues involved, is likely to undermine the potential of this tool for informed decision making. Clearly, a more critical examination of the indicators, as well as a public debate on the extent of their reliability, usefulness and acceptability to the scientific community will not only help to generate more viable indicators but also integrate scientometric studies into the mainstream of scientific activities, as well as create a demand for these studies from the scientific community and decision making bodies. With increasing size and complexity of the scientific enterprise the need for impartial and noninvasive assessment on the one hand and technology forecasting using bibliometric techniques on the other, is likely to be felt more keenly.

In summary, while trends in scientometric studies in India have changed from their being essentially an adjunct of library science to a direction where scientometricians is understood more as a tool for describing and assessing scientific activity, the aim should be to enlarge their scope and impact so that results may be utilized in a meaningful way.

## SELECTED BIBLIOGRAPHY

- Abrams, M.A., *Social Surveys and Social Action*, Heinemann, London, 1951.
- Ackoff, R.L., *The design of Social Research*, Chicago, 1967.
- Anderson, H.H. and Anderson, G.L. *An Introduction to Projective Techniques and other Devices for Understanding the Dynamics of Human Behaviour*, Prentice Hall, New York, 1951.
- Arthus, Maurice, *Philosophy of Scientific Investigation*, John Hopkins Univ. Press, Baltimore, 1943.
- Backstrom, C.H. and Hursh, G.D., *Survey Research*, North Western University Press, Evanston, III, 1963.
- Basu, A. and Garg, K.C., Changing Trends in Scientometrics/ Bibliometrics in India. In P.S. Nagpaul et al., edited Emerging Trends in Scientometrics. Allied Publishers Ltd., New Delhi, pp 18-35; 1999.
- Beveridge, W.I.B. *The Art of Scientific Investigation*, Heinemann, London, 1953.
- Bhandarkar, P.L. and Willinson T.S. *Methodology and Techniques of Social Research*. Himalaya Publishing House, Bombay, 1996.
- Blalock, H.M., Jr., and Blalock, A.B., *Methodology in Social Research*, McGraw-Hill, New York, 1968.
- Blalock, H.M., Jr., *Casual Inferences in Non-experimental Research*, University of North Carolina Press, Chapel Hill, 1964.
- Borgatta, E.F., *Sociological Methodology*, Jossey-Bass Inc., San Francisco, 1969.
- Campbell, N., *What is Science*, London, 1921.

- Campbell, D.T. and Stanley, J.C., *Experimental and Quasi-experimental Design for Research*, R and McNally, Chicago, 1966.
- Chapin, F.S., *Experimental Designs in Sociological Research*, Harper, New York, 1947
- Cochran, W. and Cox, G. *Experimental Designs*, Wiley, 1957.
- Coombs, C.H., *Theory and methods of Social Measurement*, 1953.
- Cohen, M. and E. Nagel, *An Introduction to Logic and Scientific method*, New York, 1958.
- Dewey, John, *Social Science and Social Control*, new Republic, 1931.
- Dewey, John, *Logic - The Theory of Enquiry*. New York, 1938.
- Dobry, J.T. *An Introduction to Social Research*, Stackpole, 1954.
- Dulin, R., *Theory Building*, Free Press, New York, 1969.
- Durkheim, e., *The Rules of Sociological Method*, Glencoe, 1938.
- Edwards, A.L., *Technique of Attitude Scale Construction*, Appleton Century Crofts, New York, 1957.
- Festinger, L., and Katz., D. *Research methods in Behavioural Sciences*, Dryden, 1953.
- Felsch, R.F. and Lass, A.H. *The Way to Write*, McGraw-Hill, 1957.
- Fisher, R.A., *The Design of Experiments*, Edinborough, 1949.
- Galtung, J., *Theory and Methods of Social Research*, Allen and Unwin, London,
- Garrett, H.E., Statistics in Psychology and Education. Vakils, Feffer and Simons Ltd., Bombay, 1967.
- Gibson, Q., *The Logic of social Enquiry*, London, 1958.
- Glass, R. and Glass, D.V. "Social Survey", *Chamber's Encyclopaedia* (Newnes, London), 12,667-670, 1950.
- Glock, C.Y., *Survey Research in the Social Sciences*, Russell Sage foundation, New York, 1967.

- Goode, W.G. and P.K. Hatt, *Methods in Social Research*, New York, 1952.
- Good, C.V., *On Effective Report Writing*, Philadelphia, 1957.
- Hansen, M.H., Hurwitz, W. Madow, W. Sample Survey Methods and Theory, Wiley, 1953.
- Hempel, C.G., *Fundamentals of Concept Formation in Empirical Science*, Chicago University Press, Chicago, 1952.
- Herzog, E., *Some Guidelines for Evaluative Research*, Children's Bureau, U.S. 1959.
- Hyman, H.H., *Survey Design and Analysis : Principles, Cases and Procedures*, Free Press, Glencoe, 1955.
- Kaufman, F., *Methodology of Social Science*, New York, 1944.
- Kerlinger, F.N. Foundations of Behavioural Research, Surjeet Publications, Delhi, 1978.
- Kyeberg, H.E., *Philosophy of Science - A Formal Approach*, MacMillian, New York, 1968.
- Lundberg, G.A., Social Research, New York, 1942.
- Lazarsfeld, P., "The Art of Asking Why", *National Marketing Review*, 1, 26-38, 1935.
- Lindzey, G. and Borgatta, E. *Sociometric Measurement in the Handbook of Social Psychology*, Addison-Wesley, W.S. 1954.
- Madge, John, *Origin of Scientific Sociology*, London, 1950.
- Madge, John, *The Tools of Social Science*, London, 1965.
- Merton, R.K., Social Theory and Social Structure, Free Press, New York, 1968.
- Merton, R.K., et al, *Sociology Today : Problems and Prospects*, Harper, New York, 1968.
- Moreno, J.L., *Who Shall Survive ?* New York, 1934.
- Moser, C.A., *Survey Methods in Social Investigation*, The MacMillian, New York, 1958.

- Northrop, F.S.C., *Logic of the Sciences and the Humanities*, MacMillan, N.Y., 1947.
- Pande, G.C. *Research methodology in Social Sciences*. Anmol Publication, New Delhi, 1999.
- Parsons, T., *The Structure of Social Action*, Free Press, N.Y., 1968.
- Pearson, K., *The Grammar of Science*, London, 1911.
- Polansky, Norman (Ed.), *Social Work Research*, Chicago Univ. Press, 1960.
- Popper, K.R., *The Logic of Scientific Discovery*, Basic Books Inc., New York.
- Rose, A., *Theory and method in Social Science*, University of Minnesota Press, U.S., 1954.
- Riley, M.W., *Sociological Research - A Case Approach*, New York, 1963.
- Seltiz, C., et al, *Research Methods in Social Relations*, New York, 1959.
- Stouffer, S.A., *Social Research to Test Ideas*, Free Press of Glencoe, New York, 1962.
- Suvillian, J.N.W., *The Limitations of Science*, New York, 1949.
- Theobald, D.W., *An Introduction to the Philosophy of Science*, Methuen. London, 1968.
- Torgerson, W.S., *Theory and Method of Scaling*. John Wiley and Sons. New York, 1958.
- Weber, Max, *Methodology of Social Sciences*, Glencoe, New York, 1949.
- Young, P.V., *Scientific Social Surveys and Research*, New York, 1949.
- Znaniecki, F., *The Method of Sociology*, New York, 1934.