

SCHOOL OF SCIENCE AND TECHNOLOGY

**INTRODUCTION TO NURSING RESEARCH
AND STATISTICS**

BSN 5321



BANGLADESH OPEN UNIVERSITY

SCHOOL OF SCIENCE AND TECHNOLOGY
INTRODUCTION TO NURSING RESEARCH
AND STATISTICS

BSN 5321

Course Development Team

Writers

Professor M. A. Razzaque

Dr. K. B. Sarwar

Dr. Farida Easmin Shelley

Md. Abdul Mojid Mondol

Editor

Professor K. J. A. Chisty

Abu Sayeed Md. Ripon Rouf

Program Co-ordinator

Dr. Farida Easmin Shelley

Assistant Professor

School of Science and Technology

Bangladesh Open University

Over All Supervision

Professor Khawja Jakaria Ahmad Chisty

Dean

School of Science and Technology

Bangladesh Open University



BANGLADESH OPEN UNIVERSITY

INTRODUCTION TO NURSING RESEARCH AND STATISTICS

BSN 5321

Published by:

Publishing, Printing and Distribution Division
Bangladesh Open University, Gazipur-1705

© School of Science and Technology
Bangladesh Open University

January 2006

Computer Composed and DTP Layout by:

Md. Jakir Hossin

Cover Designed by: Md. Monirul Islam

Cover Graphics: Abdul Malek

Printed by: Isamoti Offset Press

1/1, Hazi Para, Rampura, Dhaka

ISBN 984-34-4023-8

*All rights reserved. No Part of this book may be reproduced
in any form without prior permission of the copyright holder*



**SCHOOL OF SCIENCE AND TECHNOLOGY
BANGLADESH OPEN UNIVERSITY**

Contents

Introduction to Nursing Research and Statistics

Unit 1 : Introduction to Nursing Research

Lesson 1	: Nursing Research and its Historical Development	1
Lesson 2	: Ethical Considerations in Nursing Research	4
Lesson 3	: Methods and Types of Research	6
Lesson 4	: Steps of Research Process	11

Unit 2 : Conceptual Phase

Lesson 1	: Identification of Problem and Need for Study	13
Lesson 2	: Literature Review	18
Lesson 3	: Statement of the Problem	24
Lesson 4	: Purpose of Study and Development of Objectives	29

Unit 3 : Empirical Phase

Lesson 1	: Defining Variables, Formulation of Hypothesis	35
Lesson 2	: Development of Research Design, Tools and Methods of Data Collection	41
Lesson 3	: Population and Sample, Methods of Samples Selection	56
Lesson 4	: Methodology and Action Plan for Data Collection	69

Unit 4 : Interpretive Phase

Lesson 1	: Basic Statistical Concept and Common Terms	81
Lesson 2	: Processing, Analysis and Interpretation of Data	98
Lesson 3	: Writing the Research Report	109

Unit 2: Conceptual Phase

Lesson 1: Identification of Problem and Need for Study

1.1. Learning Objectives



On completion of this lesson you will be able to-

- identify for a research problem
- the problem analysis.



1.2. Introduction

There are countless problem in the world, such as, in your workplace; in your administration; in your service areas and so on. Finding a problem is not hard, but identify one for the purpose of research is really a tough job. One of the most important tasks of research is to identify and define clearly the problem you wish to study.

If you are uncertain about the problem of research, if you are not settled in your mind about what you want to examine, you might be sure about your teachers who would read your proposal would also be uncertain.

A well-defined problem leads naturally to the statement of research objectives, to the hypothesis, to a definition of key-variables to a selection of a methodology for measuring the variable. A poorly defined research problem always leads to confusion. All research is set in motion by the existence of a problem.

A problem is a perceived difficulty a feeling of discomfort with the way things are a discrepancy between what someone believes should be and what is.

All problems do not require research. A potential research situation arises when 3 conditions exist:

1. A perceived discrepancy between what is and what should be.
2. A question about why the discrepancy exists.
3. At least two possible and plausible answers to the question.

Research Process Conceptual Phase

The last point is very important. If there is only one possible and plausible answer to the question about the discrepancy, then a research situation does not exist. Consider the example below:

A. Example of a Non research Problem

Problem Situation: A recent survey in district A found that 1,000 women were continuous users of contraceptive pills. But last month's survey statistics indicate that none of those women are using contraceptive pills.

Discrepancy: All 1,000 women should use contraceptive pills, but all 1,000 women are not using contraceptive pills.

Problem Question: What factor or factors are responsible for 1,000 women discontinuing their use of contraceptive pills?

Answer: A monsoon flood has prevented all new supplies of pills reaching district A, and all old supplies have been exhausted.

In this above example, a problem situation exists, but the reason for the problem is already known. Therefore, assuming that all the facts correct, there is no reason to conduct research on the factors associated with pill discontinuation among the 1,000 women. On the other hand, there may very well be a need to conduct research on the question of why the supply and logistics system is incapable of providing contraceptives to women during the monsoon.

Study the next example.

B. Example of a Research Problem

Problem Situation: District B is always flooded during the monsoon season. Recognizing this problem, the national family planning program established a new supply logistics system for the district. Each pill user is given a four-month supply before the monsoon begins. During the monsoon, small motorboats are available to transport new supplies to selected distribution centers accessible to village-level family planning workers. Despite those new measures, this year services statistics indicated that there are no pill supplies in District B.

Discrepancy: The new logistics system should be able to assure a continuous supply of pills, but this year there are no supplies.

Problem Question: Why has the new supply logistics system been incapable of delivering contraceptive pills to users?

Possible Answer

1. An order for new pill supplies was not placed in time before the monsoon rains.
2. The riverboats used to transport the supplies are out of order.
3. Field-workers were not told about the new system and failed to give users a four-month supply of pills before the monsoon.

In this example, there are several possible and plausible reasons for the problem situation. One or more of these reasons might be correct. Therefore, this is a potential research situation.

In some situations, it is relatively easy to identify the problem, to define it, to hypothesize the reasons for it, and to conduct research to determine which reason is correct or more nearly correct. The reasons for the supply and logistics problem in the above example could probably be determined fairly easily and certainly would not require an expensive research study. Other problems, such as the one in the next example, are not so easy to identify or to study:

C. Example of another Research Problem

Problem Situation: A recent family planning survey revealed great differences between villages in the rate of contraceptive prevalence. Despite the fact that all villages receive the same level of health and family planning services, some villages have a prevalence rate as high as 80 percent, while others have a rate as low as 6 percent.

Discrepancy: All villages should have approximately the same rate of contraceptive prevalence, but in fact there is great variation between villages.

Problem Question: What factors are responsible for the aerial variation in contraceptive prevalence rates?

Possible Answers

1. Village differs in their socioeconomic environments. Some villages are agricultural; some are fishing communities. Some villages are Hindu; others are predominantly Muslim or Buddhist. Some villages have access to market in towns; others do not. Some villages have schools, health clinics, electricity, and a good water supply; others do not have these facilities. These socio-economic differences affect the level of contraceptive practice.

Research Process Conceptual Phase

2. Village differs in institutional support for contraceptive acceptance. In some villages, local influential strongly support the national family planning program. In other villages, they do not support it. In some villages, there are active Mothers' Clubs that support family planning. In other village, there are no Mothers' Clubs. These differences in institutional support for family planning affect the level of contraceptive practices.
3. Village-level health and family planning workers differ in their effectiveness. Some workers are highly motivated and very active in their assigned areas. Other workers are less motivated and less active. These differences in worker effectiveness affect the level of contraceptive practice.

While the problem situation presented above is fairly clear, the reasons for the problem are complex. Three reasons have been given, but it is likely many more could be stated. In situation like this, the researcher must devote considerable time and attention to identify problem situation. The aim is to focus the research on the most important aspects of the problem.

1.7. Exercise

Read the following problem and identify discrepancy and possible answers:

During their periodic visits to clients, the supervisors in a rural, community-based family planning program observe that many users of oral contraceptives are generally poorly informed about the appropriate use of the pill. For examples, 58 percent of the women consulted had taken the pill incorrectly during the past month. Some had waited either too long or not long enough after the menstrual cycle; others did not know what to do when they missed taking a pill.



1.7. Exercise

1.7.1. Multiple choice questions

Tick (✓) the correct answers

1. A problem is situation that has
 - a. Single answer
 - b. Multiple answers
 - c. No answer
 - d. All of the above.

2. Discrepancy means
 - a. Gap between what is and what should be
 - b. Problem
 - c. Situation analysis
 - d. None of the above.

3. Research is a complex process that needs
 - a. In-depth search
 - b. Continuous inquiry
 - c. Comparison of related answers
 - d. All of the above.

1.7.2. Short questions

1. Define research.
2. How many condition is needed for a research define it?

Lesson 2: Literature Review

2.1. Learning Objectives



On completion of this lesson you will be able to-

- what is literature review
- sources of literature review
- how to do a good literature review.



2.2. Introduction

To say literature, in research, we mean the work(s) done before-published, unpublished, or in raw form- in the same or related field. The procedure or methodology, the finding or results and the recommendations can be mentioned, judged and or compared with the new one. It is compulsory for a research worker writing a thesis or dissertation quoting the other work done previously put forward with the synopsis before the research work undertake.

At this point, researcher should pass through extensive literature survey connected with the problem s/he identified before. For this, the abstracting and indexing journals and published or unpublished bibliographic are the first place to go to.

All related data, interpretation, results should be noted in a scientific fashion that can be used as reference or for comparison the earlier studies, if any, which are similar to the study in hand should be carefully studied. A good library will be a great help to the researcher in this stage.

2.3. Definition of Literature Review

The review of literature is defined as a broad, comprehensive in depth, systematic and critical review of scholarly publications, unpublished print materials, audio-visual matter and personal communications.

2.4. Objectives

The main objectives of the review of literature are as follows:

1. To determine what is known and not known about a subject, concepts or problem
2. To determine gaps, consistencies, and inconsistencies in the literature about a subject, concept or problem

3. To discover unanswered questions about a subject, concepts or problems
4. To describe the strength and weaknesses of designs/methods of inquiry and/or instruments used in earlier works
5. To discover conceptual traditions used to examine problems
6. To generate useful research questions or projects/activities for the discipline
7. To determine the appropriate research design/method (instruments, data collection and analysis methods) for answering research question
8. To determines the need for replication of well-designed study or refinements of a study
9. To promotes developments of protocols and policies related to nursing, practice, (service, administration, education and research)
10. To uncover a new practice intervention of gain support for changing a practice intervention.

2.5. Sources of the Literature Review

The types of information sources for a review of literature are conceptual and data based literature. The common sources of both these literatures are books, journal articles, abstracts, critique reviews, abstracts published in conference proceedings, professional and governmental reports, and unpublished doctoral dissertations.

The kinds of information available in written documents can be categorized into 5 broad classes.

1. Facts, findings or results
2. Theory
3. Research procedure or methods
4. Opinions, points of view or personal commentaries
5. Anecdotes or impression on a particular event or situation.

The references can be categorized has being either *primary* or *secondary* sources.

Primary Sources

A primary sources is written by a person who developed the theory or conducted the research or is the description of an investigation written by

Research Process Conceptual Phase

the person who conducted it. Most primary sources are found in published literature. e.g. Nursing research article.

A credible literature review reflects the use of mainly primary sources.

Secondary Sources

A secondary source is written by a person other than the individual who developed the theory or conducted research. Otherwise, known, as it is description of study or studies prepared by someone other than the original researcher. Often secondary source represents a response to, or a summary and critique of a theorist or researcher's work. The secondary sources may be used when unavailability of primary sources, and if we want to know different ways of looking at an issue or problem.

The following recommendation to increase critical evaluation skills when reading above primary and secondary sources of literature-

- Read the primary sources of a study or theory- not just secondary source
- Seek assistance about the critique researcher
- Read secondary sources from referred or pre-reviewed journals
- Discuss your response to secondary source articles
- Variables those were defined.

In searching literature, the researcher should note certain important elements given below:

- Reports of studies closely related problems that have been investigate
- Design of the study
- Population of the study
- Faults that could have been avoided
- Recommendation for further research.

2.6. Some Selected Sources for Review of Literature

1. Journals
2. Abstracts and excerpts
3. Bibliography
4. Encyclopedias

5. Handbooks
6. Indexes
7. Inventories etc.

Example of some computerized database: CD-ROM and UN-LIM

1. CINAHL- Cumulative Index to Nursing and Allied health literature
2. MEDLINE- Standard-Medical Literature Analyses and Retrieval system on line (MEDLAR)
3. PSYCHLIT- Psychology Literature
4. ERIC- Educational Research Information center.

2.7. Tips on How to Do a Good Literature Review

In order to get most relevant and appropriate material for study the researcher should learn the correct techniques of literature reviews. Here are a few tips for the beginner-

- Be selective
- SCAN the summaries and abstracts of the articles first and then “LIFT of LEAVE” the journal depending on whether the material meets your needs.
- Take prompt notes. Do not postpone note taking.
- Use small index cards to take notes and write journal references on the reverse side.
- Use one card for each reference.
- Arrange and store the cards carefully (you can use an empty 1 litre ice cream box for this purpose).

Example of an Index Card

Serial No _____
Author(s) _____
Title of article _____
Journal Title _____
Vol No Month Year Pages
publishers, place of publication, date, year (for Books)

Library where found Call Number
Source of bibliographic information:

Research Process Conceptual Phase

(On the reverse, note the content of the article that is related to the research problem in hand)

For a serious student we advise that cards be prepared in duplicate and kept in two packs (bundles or files). One where the cards are arranged in order of their serial number, the other where these are in alphabetical order. Both these collections will have their uses and save hours of search for a forgotten reference.

The literature review is best done and presented in the form of an inverted pyramid, the broad, general, global areas first and specific later.

2.8. When to Review

Beginners in research usually postpone review of literature till the time of report writing. This is a common mistake. Literature review is a continuous and, unending process. A researcher needs to cultivate the habit of reading important literature. Expressed in arbitrary proportions, the beginners should allocate time to this activity approximately as follows-

Before undertaking research work	60-75%
During the research work	10-20%
After completing data analysis but before writing the research report.	20-25%



2.9. Exercise

2.9.1. Multiple choice questions

Tick (✓) the correct answers

1. In research, literature means
 - a. Poetry
 - b. Prose
 - c. Related research
 - d. None of the above.

2. Secondary source does not include
 - a. Book
 - b. Thesis
 - c. Magazine
 - d. Newspaper

3. Medline is a
 - a. Internet source
 - b. Virtual source
 - c. Distance source
 - d. All of the above.

2.9.2. Short questions

1. Why literature review is important for every researcher?
2. What are the differences between primary and secondary sources of literature?
3. How literature review can help a research?

Lesson 3: Statement of the Problem

3.1. Learning Objectives



On completion of this lesson you will be able to-

- analysis of a problem.



3.2. Introduction

You have learned how to identify a problem and how to define problem situation. It is now necessary to understand the way to describe the problem.

The objective should specify WHO will do, How much of what, to whom when and when for WHAT purpose.

Research often is expensive and time consuming. Ask yourself, why is the problem I wish to study important? Can you judge your selection of the research problem? Can you convince other friends, colleagues or tutors about the importance of the problem?

3.3. Problem Identification

Identifying a problem situation is the first essential step in designing a research proposal, but a process of problem identification must then follow it.

The research problem identified must now be defined in terms of its occurrence, intensity, distribution, and other measures for which data are already available. The aim is to determine all that is currently known about the problem and why it exists. While it is always possible to guess why a problem exists, guesses are often wrong and usually they do not provide a firm basis for designing a research study. A far better way to define a problem situation is to review relevant literature, examine current service statistics, seek educated opinions from persons concerned with the problem, and obtain probable reasons for the problem from social, economic, or health theory. A careful review of existing sources of information on a problem and an epidemiologic diagnosis help the investigator determine:

1. **Incidence and Prevalence:** An epidemiologic diagnosis should always be made of problems related to health and family planning. In other words, how widespread is the problem? What is its distribution? How often does it occur? An epidemiologic diagnosis will help establish the parameters of the problem.
2. **Geographic Areas Affected:** It is important to know if there are particular geographic areas affected by the problem. Does the problem

generally occur in rural areas only? Does it also affect cities? Is the problem restricted to mountain areas, coastal areas, or island areas?

3. **Characteristics of Population Groups:** Are there special population groups affected by the problem, such as mothers, children, teenagers, newly married couples, or women over age 35?
4. **Probable Reasons for the Problem:** A review of information on a problem should suggest a number of probable reasons why the problem exists. What is the current thinking about the reasons for the problems? Is there general agreement among many people about the reasons, or are there many different, conflicting views?
5. **Possible Solutions:** Many projects and programs may have been directed at the problem in an attempt to overcome it. What types of solutions have been tried in the past? How successful have past efforts been? What approaches to solving the problem seem to work? What approaches seem not to work?
6. **Unanswered Question:** From the review of information on the problem, what seem to be the unanswered question about the problem? What aspects of the problems need to be research further?

Reviewing what is already known about a problem situation is an essential part of the research process. A good review of information will suggest the social, economic, political, and health importance of the problem. It will help to narrow the focus of the proposed research. It will indicate the major theoretical concepts and operational variables other researchers have considered important. It will suggest possible research hypotheses that need to be tested. Finally, it will help the investigator avoid reinventing the wheel, or in other words, conducting research on a problem that has already been researched countless times in the past with fairly consistent findings from each study.

3.4. Research Problem Identification

1. Although substantial progress toward reducing fertility rates was made during the last decade in several Southeast Asian countries, significant program constraints still exist and threaten further fertility declines. In Indonesia, the relatively high contraceptive prevalence rate on Java and Bali of 42 percent must be viewed against the relatively low rate of 17 percent in the Outer Island-I area and only 6 percent in the Outer Island-II area. In the Philippines, only 25 percent of all contraceptive practice involves the use of modern methods. In Malaysia, new sterilization acceptors constitute less than 6 percent of all new acceptors and new IUD acceptors less than 3 percent. In Thailand, the vasectomy program is lagging despite some remarkable acceptance rates demonstrated when private sector teams have taken services directly to villages.

Research Process Conceptual Phase

2. Most program administrators within Asia are aware that a new level of effort and direction will be needed in order to sustain the trends established in the 1970s. In Indonesia, attention needs to be given not only to the Outer Island areas but also to the urban areas, particularly Jakarta. Although Jakarta has a health and family planning service delivery system designed to cover all areas of the city, the contraceptive prevalence rate is only 21 percent of currently married women, or in other words, approximately half of the rate for all of Java and Bali. Other large urban areas in the country also lag behind the rural areas.
3. While the factors responsible for the difference in contraceptive prevalence rates between urban and rural areas in Indonesia are not known, it has been suggested that urban areas lack the traditional structures of community organization and leadership that help support the family planning program in rural areas. Another possible reason for the difference might be lack of awareness, particularly among new migrants to urban areas, of sources of contraceptive supply. Many of the new migrants have low levels of education, do not have access to sources of information, and are unaware of the location of health and family planning service centers. Also, family planning city workers often find it difficult to provide continuous services to new migrants at their homes since these people tend to shift their residence periodically from one area of the city to another. It is likely that; among acceptors, these shifts result in discontinuation, or at least in an interruption of contraceptive practice.
4. In the past, several attempts have been made to increase the urban rate of contraceptive prevalence by providing more accessible services to residents. In Jakarta, mobile vans have been used to bring services to outlying areas of the city. Another approach has been to distribute condoms and pills through such commercial outlets as small shops and tea stalls. While both of these approaches are useful, they also have significant drawbacks. Mobile vans are expensive to operate, and they can reach only a relatively small proportion of the population of potential acceptors. Commercial outlets are helpful for current acceptors, but they are not very effective as a means to motivate new acceptors. Also, these outlets cannot provide needed maternal and child health services, such as immunization and pre-and postnatal care.
5. What is needed in urban areas is some means of (1) reaching large numbers of current and potential acceptors and (2) providing them with a range of health and family planning services. An important unanswered question for the national family planning program is how this can best be accomplished.
6. One approach that has been suggested but so far not tried is to train a special category of urban health and family planning personnel who would contact acceptors and potential acceptors at their places of

employment. With the assistance of employers and unions, the urban health and family planning worker could establish service delivery centers at factories, government offices, hotels, construction sites, markets, and other places of employment. It is likely that this approach would be less than the use of mobile vans; it would be capable of reaching large numbers of people; and the range of services that could be provided would be greater than that provided by commercial outlets.

Comments on the Example

- a. In the first paragraph of this example, a broad problems situation is identified. Substantial progress toward reducing fertility rates in several Southeast Asian countries has been made, but continued progress is threatened by significant program constraints. Several examples of program constraints are then given. These examples serve to indicate to geographic areas affected by the problem.
- b. In the second paragraph, the problem of program constraints is narrowed down to just one country-Indonesia; also, attention is focused on just one program problem- the difference in contraceptive prevalence rates between urban and rural areas. A discrepancy is indicated. Jakarta has a health and family planning service delivery system designed to cover the entire city. This system should result in prevalence rates equal to the rate in rural areas. But the prevalence rate in Jakarta is only half the rate found in all of Java and Bali. By the end of the second paragraph, we know the specific problem situation-urban/rural differences in contraceptive practice. We know the location of the problem-Indonesia, urban areas, particularly Jakarta. We know the magnitude of the problem-21 percent prevalence in Jakarta compared with 42 for both Java and Bali. We know the specific population of interest-urban residents.
- c. In the third paragraph, several possible reasons for the problem situation are suggested. It is not known which one of these reasons is correct or more nearly correct.
- d. The fourth paragraph outlines several possible solutions to the problem of low prevalence rates in urban areas. Mobile vans have been used. Commercial outlets have been used. Each of these approaches to the problem is helpful, but each is rejected as inadequate.
- e. The fifth paragraph suggests what is needed to overcome the problem, and the sixth paragraph outlines a new and previously untested approach to overcoming the problem.

3.5. The Closing

In this, a problem can be address by its magnitude, by length and by depth measurement. Following the specific data or issue, a real situation can be described in such a way that can be indicated clearly a problem, which should be examined by the researcher herself/himself.



3.6. Exercise

3.6.1. Multiple choice questions

Tick (√) the correct answers

1. What were the possible way of contraceptive constrain in Jakarta?
 - a. One
 - b. Two
 - c. Three
 - d. Four.

2. Incidence and prevalence is same, if it is counted in
 - a. Same period
 - b. Long period
 - c. Form history
 - d. None of the above.

3. What is your opinion on research does it-
 - a. Expensive
 - b. Time consuming
 - c. Expensive and time consuming
 - d. Non of the above.

3.6.2. Short questions

1. What do you mean by problem statement process?
2. Why problem statement is important for research?
3. Give three examples of problem statement.

Lesson 4: Purpose of Study and Development of Objectives

4.1. Learning Objectives



On completion of this lesson you will be able to-

- how to highlight the purpose of study
- what is objective, hypothesis and variable
- how to address these.



4.2. Objectives and Hypotheses

After completion of identification of the problem, literature search and statement of the problem, how the time to fix the purpose with definite objectives and hypothesis.

The objectives and hypotheses of a research study should flow logically from the earlier sections identifying the problem situation, defining the parameters of the problem, and justifying its importance. In this section, the research purpose is narrowed and focused. Specific objectives are written that describe the expected contributions and outcomes arising from the study and the variables that with the measured. Also, specific, testable hypotheses are formulated that give the relationship between variables.

4.3. Ultimate Objectives

Most research studies have a ultimate objectives. Ultimate objectives help to relate the proposed research to broad social, economic, and health concerns. The ultimate objectives of many health and family planning studies is to contribute in some way toward reducing morbidity, mortality, or fertility. This of course is a rather unrealistic objective. It is unlikely that any single study will do much to reduce morbidity, mortality, or fertility.

Objectives should be SMART:
S- Specific
M- Measurable
A- Accessible
R- Reasonable
T- Time-bound

Example of Ultimate Objectives

1. **Study of Areal Variation in Contraceptive Practice in Bangladesh:** The ultimate objective of this study is to help program administrators to design new educational strategies for motivating couples in villages with low contraceptive prevalence.
2. **Study in Burkina Faso to Test an Approach to Providing Family Planning Services to Urban Resident:** The ultimate objective of this study is to develop a cost effective and sustainable model for

providing quality health and family planning education and services to urban residents.

3. **An Experimental Study to Integrate Community based Family planning Education and Service with Primary Health Care in Two Rural Areas of Cameroon:** The ultimate objective of this study is to develop a rural, integrated, cost-effective family planning and health services system which has a measurable impact on decreasing infant and maternal morbidity and mortality.
4. **Survey Study of the Family Planning Knowledge, Attitudes and Practices of Community Influential in Sri Lanka:** The ultimate objective of this study is to increase the involvement of community influential in health and family planning program activities.

4.4. Immediate Objectives

In contrast to ultimate objectives that state what is expected to happen immediate objectives state what will happen. Immediate objectives relate directly to the research problem situation. An immediate objective represents a promise by the investigator that certain activities will take place and specific variables will be examined.

Whenever possible, and certainly in all evaluation studies, the immediate objectives should be stated in behavioral terms- in other words, in terms of what will be done. The objective should specify: Who will do, how much of what, to whom, where, and for what purpose.

4.5. Hypothesis, Variables and Their Relationship

When writing hypothesis statements, it is important to keep in mind the distinction between independent and dependent variables. An independent variable causes or determines or influences the dependent variable. The basic relationship between these two types of variables can be represented as in Fig. 1. This model shows a direct relationship. In other words, whenever the independent variable changes, there is a change in dependent variable. The dependent variable depends on the independent variable.

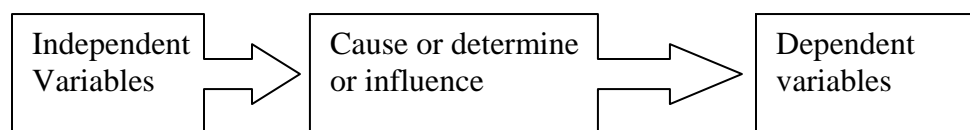


Fig. 1: Direct relationship between independent and dependant variable.

In most family planning operation research studies, it is unusual to find a direct relationship of this type. Almost always, the relationship is indirect. The independent variable acts upon the dependent variable through

intervening variables. These intervening variables serve either to increase or to decrease the effect the independent variable has on the dependant variable. Intervening variables are sometimes referred to as control variables. Many, if not most, family planning research studies are based on a model that includes intervening variables.

For example, in an effort to increase contraceptive acceptance (the dependent variable), a family planning program might initiate a large information, education, and communication (IEC) program (the independent variable). Alone and by itself, an IEC program cannot increase contraceptives use. The program must act on and operate through a set of intervening variables, which in turn cause or determine or influence contraceptive use. There may be many of this intervening knowledge about family planning, their attitudes toward the use of birth control methods, and their beliefs about family size. A possible research model for an evaluation study of the effects of an IEC program is shown in Fig. 2.

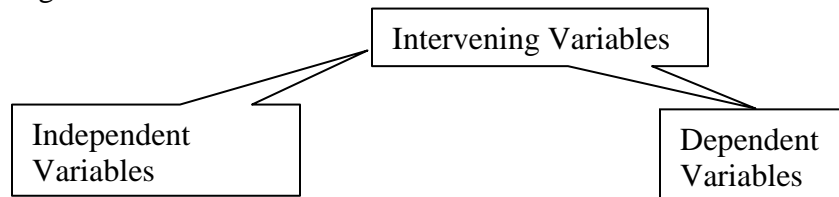


Fig. 2: Intervening variables

Most research studies will examine many independent variables and many intervening tables but only a few dependent variables. In writing study hypotheses, always think in terms the expected relationship between variables. Think first about the central problem your study address (the dependent variable). Next consider what factor or factors (the independent tables) might cause, determine, or influence the problem situation. Finally, ask yourself if the between the independent and dependent variables is direct or indirect through a set intervening variables.

Examples of Hypotheses

1. Villages with many modern characteristics will have a higher rate of contraceptive practice than villages with few modern characteristics.
2. Contraceptive practice will be higher in villages where the family planning field-worker is viewed as credible than in villages where the field-worker is not viewed as credible.
3. The work performance of health and family planning staff who have received a five week field-based training course will be higher than the work performance of similar staff who have received a four-week, classroom-based training course.

Research Process Conceptual Phase

4. Higher levels of field-worker performances will lead to higher rates of contraceptive acceptance among eligible couples.
5. An information and education program using radio, television, and printed material will increase people's knowledge about family planning.
6. People with higher levels of family planning knowledge will be more likely to be users of contraceptive methods than people who are not as knowledgeable about family planning.

Comments on the Example

Note that in each hypothesis there is a statement of an expected relationship between two or more variables. In this first hypothesis, the expected relationship is between modern characteristics of villages (the independent variable) and contraceptive practice (the dependent variable). In the second hypothesis, the relationship is between the credibility of field-workers (the independent variable) and contraceptive practice (the dependent variable). The third and fourth hypotheses go together. In the third hypothesis, the relationship is between a five week field-based training program (the independent variable) and field-worker performance (the intervening but also dependent variable). In the fourth hypothesis, field-worker performance acts as an independent variable and contraceptive acceptance rates becomes the dependent variable. Then together, the model for those two hypotheses would look like Fig.3 Similarly, hypotheses five and six go together. Family planning knowledge is the dependent variable in hypothesis five, but becomes the independent variable in hypothesis six.

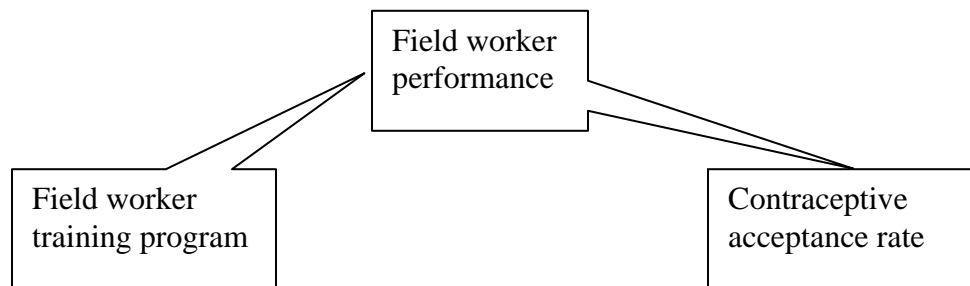


Fig. 3: A model for two hypotheses about the effect of a field worker-training program on contraceptive acceptance rate

To summarize, all research studies should include a statement of (1) ultimate objectives, (2) immediate objectives, and (3) hypotheses (except in the case of purely exploratory or descriptive studies). Each of these statements serves a different purpose:

1. Ultimate objectives state the anticipates (hoped for) contributions of the study.

2. Immediate objectives state what will be done in the study.
3. Hypotheses state the expected relationship between two or more variables.

What to do – Writing Ultimate Objectives, Immediate Objectives, and Hypotheses

List the major variables of your study under the headings: Independent variables and dependent variables. Intervening variables can be listed under both headings. The dependent variables should relate directly to your problem statement.

Review the list of dependent variables, and then write a statement of ultimate objectives that relates to them. Ask yourself the following question: If I knew the factors that caused, determined, or influenced the dependent variable(s), how would this knowledge help policy-makers, program administrators, or others? How will the results from the study help improve other words, what are the anticipated contributions of the study?

Now write the immediate objectives for your study. Focus your attention on the specific actions that you will perform. Exactly what do you plan to do? What variables do you plan to measure? Write behavioral objectives that answer the following questions:

- a. Who will do?
- b. How much of what?
- c. To whom?
- d. When?
- e. Where?
- f. For what purpose?

Finally, write the hypotheses that your study will test. Write hypothesis statements in positive, not negative, terms. Write a hypothesis for each major relationship that you expect to test in your study.



4.12. Exercise

4.12.1. Short questions

1. Define objectives.
2. What is the difference between hypothesis and objectives?
3. How to write objectives?

Unit 3: Empirical Phase

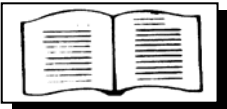
Lesson I: Defining Variables, Formulation of Hypothesis

1.1. Learning Objectives



On completion of this lesson you will be able to-

- conceptual and operational definitions of a variable
- different types of variables
- importance and uses of different types of variables
- the meaning of research and hypothesis
- objectives of research
- different types of hypothesis.



1.2. Variables

Variables are qualities, properties, or characteristics of persons, things, or situations that changes or vary. Variables are also concepts at different levels of abstraction that are concisely defined to promote their measurement or manipulation within a study. The concepts like weight, height, and income are all examples of variables.

1.3. Conceptual and Operational Definitions of Variables

A variable is operationalized in a study by the development of conceptual and operational definitions. A **conceptual definition** provides the theoretical meaning of a variable and is often derived from a theorist's definition of a related concept.

An **operational definition** is derived from a set of procedures or progressive act that a researcher performs to receive sensory impressions (such as sound, visual, or tactile impressions) that indicate the existence or degree of existence of a variable. Operational definitions need to be independent of time and setting so that variables can be investigated at different times and in different setting using the same operational definitions. An operational definition is developed so that a variable can be measured or manipulated in a concrete situation, and the knowledge gained from studying the variable will increase the understanding of the theoretical concept that this variable represents.

1.4. Types of Variables

Variables have been classified into a variety of types to explain their use in research. Some variables are manipulated others are controlled. Some

Empirical Phase

variables are identified but not measured; others are measured with refined measurement devices. The types of variables presented in this section include continuous, discrete, independent, dependent, research, extraneous, and demographic.

Continuous variables: These variables can take any value within a range. Our age is an example of continuous variable.

Discrete variables: If the variables can be expressed in integer values, they are non-continuous or in statistical language “discrete variables”. The number of books is a discrete variable.

Independent Variable: An independent variable is a stimulus or activity that is manipulated or varied by the researcher to create an effect on the dependent variable. The independent variable is also called a treatment or experimental variable.

Dependent Variables: A dependent variable is the response, behavior, or outcome that the researcher wants to predict or explain. Changes in the dependent variable are presumed to be caused by the independent variable. If we say height depends upon age, then height is the dependent variable and age is the independent variable.

Research Variable: Research variables or concepts are the qualities, properties, or characteristics identified in the research purpose and objectives or questions that are observed or measured in a study. Research variables or concepts are used when the intent of the study is to observe or measure variables, as they exist in a natural setting without the implementation of a treatment.

The extraneous variables that are not recognized until the study is in process, or are recognized before the study is initiated but cannot be controlled, are referred to as **confounding variables**. Sometimes extraneous variables can be measured during the study and controlled statistically during analysis. However, if these extraneous variables cannot be controlled or measured, this hinders the interpretation of findings.

Demographic Variables: Demographic variables are characteristics or attributes of the subjects that are collected and analyzed to describe the study sample. Some common demographic variables are age, education, gender, ethnic origin (race), marital status, income, job classification, and medical diagnosis (es). When a study is completed, the demographic data are analyzed to provide a picture of the sample and are called sample characteristics.

1.5. Research

A hypothesis is a formal statement of the expected relationship (s) between two or more variables in a specified population. The hypothesis translates the research problem and purpose into a clear explanation or prediction of the expected results or outcomes of the study. A clearly stated hypothesis includes the variables to be manipulated or measured, identifies the population to be examined, and indicates the proposed outcomes for the study. Hypotheses also influence the study design, sampling technique, data collection and analysis methods, and interpretation of findings. A hypothesis states what we are looking for and it is a proposition, which can be put to a test to determine its validity.

Characteristics of hypothesis: Hypothesis must possess the following characteristics:

- i. Hypothesis should be clear and precise. If the hypothesis is not clear and precise, the inferences drawn on its basis cannot be taken as reliable.
- ii. Hypothesis should be capable of being tested. In a swamp of untestable hypotheses, many a time the research programmes have bogged down. Researcher may do some prior study in order to make hypothesis a testable one. A hypothesis is testable if other deductions can be made from it, which, in turn, can be confirmed or disproved by observation.
- iii. Hypotheses should state relationship between variables, if it happens to be a relational hypothesis.
- iv. Hypothesis should be limited in scope and must be specific. A researcher must remember that narrower hypotheses are generally more testable and s/he should develop such hypotheses.
- v. Hypothesis should be stated as far as possible in most simple terms so that the same is easily understandable by all concerned. But one must remember that simplicity of hypothesis has nothing to do with its significance.
- vi. Hypothesis should be amenable to test within a reasonable time. One should not use even an excellent hypothesis; if the same cannot be tested in reasonable time for one cannot spend a lifetime collecting data to test it.
- vii. Hypothesis must explain the facts that gave rise to the need for explanation. This means that by using the hypothesis plus other known and accepted generalizations, one should be able to deduce the original problem condition. Thus hypothesis must actually explain what it claims to explain; it should have empirical reference.

1.6. Null versus Research Hypotheses

The null hypothesis symbolised as (H_0), also referred to as a statistical hypothesis, and is used for statistical testing and for interpreting statistical outcomes. Even if the null hypotheses in not stated, it is implied, because it is the converse of the research hypothesis. Some researchers state the null hypothesis because it is more easily interpreted based on the results of statistical analyses. The null hypothesis is also used when the researcher believes there is no relationship between two variables and when there is inadequate theoretical or empirical information to state a research hypothesis. A null hypothesis can be simple or complex and associative or causal.

A research hypothesis is the alternative hypothesis (symbolized as H_1 or H_a) to the null hypothesis and states that there is a relationship between two or more variables. Research hypothesis can be simple or complex, nondirectional or directional, and associative or causal.

1.7. Testable Hypothesis

A testable hypothesis is clearly stated, predicting a relationship between or more variables. Hypotheses are clearer without the phrase “There is no significant difference” because the level of significance is only a statistical technique applied to sample data. The value of a hypothesis is ultimately derived from whether it is testable in the real world. A testable hypothesis is one that contains variables that are measurable or able to be manipulated. Thus, the independent variable must be clearly defined so that it can be implemented precisely and consistently as a treatment in the study. The dependent variable must be precisely defined to indicate how it will be accurately measured. A testable hypothesis also needs to predict a relationship that can be “supported” or “not supported” based on data collection and analysis.

Examples: Different types of hypotheses with variables identified

- i. “There will be (is) no difference between men and women in knowledge about HIV transmission routes (sexual, needle sharing, casual)”. This hypothesis is null, associative, simple, and nondirectional, with two variables, gender and knowledge about HIV transmission routes.
- ii. “There will be (is) no difference between men and women in knowledge about the effectiveness of measures to prevent sexual transmission of HIV”. This hypothesis is null, simple, associative, and nondirectional, with two variables, gender and knowledge about effectiveness of measures to prevent sexual transmission of HIV.

- iii. “Men and women will differ with respect to reported frequency and type of behaviors that could lead to the transmission of HIV”. This hypothesis is research, associative, complex, and nondirectional. The variables are gender, frequency of behaviors, and types of behaviors. This hypothesis is nondirectional because there is no indication of how men and women will differ and associative because a relationship is stated with no indication of cause and effect.
- iv. “Men and women will differ with respect to reported level of comfort and safer sexual practices”. This hypothesis is research, complex, associative, and nondirectional. The variables are gender, level of comfort, and safer sexual practices. The hypothesis is nondirectional, with no indication of how men and women will differ, and associative because the relationship does not indicate cause and effect between variables.



1.8. Exercise

1.8.1. Multiple-choice questions

1. Which one is a variable?
 - a. Man
 - b. Woman
 - c. Nurse
 - d. Height.
2. Number of books is a
 - a. Continuous variable
 - b. Discrete variable
 - c. Environmental variable
 - d. Demographic variable.
3. A hypothesis should be
 - a. Ambiguous
 - b. Clear and precise
 - c. Non formal statement
 - d. Untestable.
4. Hypothesis can be described using
 - a. Two categories
 - b. Three categories
 - c. Four categories
 - d. Five categories.

Empirical Phase

1.8.2. Short questions

1. Define a variable and give three examples.
2. What do you mean by a hypothesis?
3. What are directional and non-directional hypotheses?
4. Define simple and complex hypotheses.

1.8.3. Analytical questions

1. Clarify operational and conceptual definitions of a variable.
2. Describe the characteristics of a hypothesis.
3. Explain Null versus Research hypotheses.

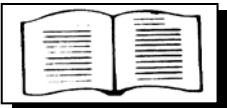
Lesson 2: Development of Research Design, Tools and Methods of Data Collection

2.1. Learning Objectives



On completion of this lesson you will be able to-

- the purpose of a research design
- how to develop a research design
- different types of designs
- different concepts for selecting a design
- four levels of measurement of data collection
- the common tools and methods of data collection used in nursing research
- the process of data collection.



2.2. A Research Design

A research design is a blueprint for conducting a study that maximizes control over factors that could interfere with the validity of the findings. Elements control to the study design include the presence or absence of a treatment, number of groups in the sample, number and timing of measurements to be performed, sampling method, time frame of data collection, planned comparisons, and control of extraneous variables.

In fact the research design is the conceptual structure within which research is conducted, it constitutes the blue print for the collection, measurement and analysis of data. In brief, research design must, at least contain-(i) a clear statement of the research problem; (ii) procedures and techniques to be used for gathering information; (iii) The population to be studied; and (iv) Methods to be used in processing and analyzing data.

The purpose of design is to maximize the possibility of obtaining valid answers to research questions or hypotheses. In most studies comparisons are the basis of obtaining valid answers. A good design provides the subjects, the settings and the protocol within which these comparisons can be clearly examined. Critiquing a design involves examining the study's environment, sample, treatment, and measurement.

One may split the overall research design into the following parts:

- i. **The sampling design** which deals with the method of selecting items to be observed for the given study;
- ii. **The observational design** which relates to the conditions under which the observations are to be made;

Empirical Phase

- iii. **The statistical design** which concerns with the question of how many items are to be observed and how the information and data gathered are to be analyzed; and
- iv. **The operational design**, which concerns with the techniques by which the procedures specified in the sampling, statistical and observational designs can be carried out.

To develop a research design we must have to answer to the following questions:

- i. What is the study about?
- ii. Why is the study being made?
- iii. Where will the study be carried out?
- iv. What type of data is required?
- v. Where can the required data be found?
- vi. What periods of time will the study include?
- vii. What will be the sample design?
- viii. What techniques of data collection will be used?
- ix. How will the data be analyzed?
- x. In what style will the report be prepared?

So for development of research design we should have to:

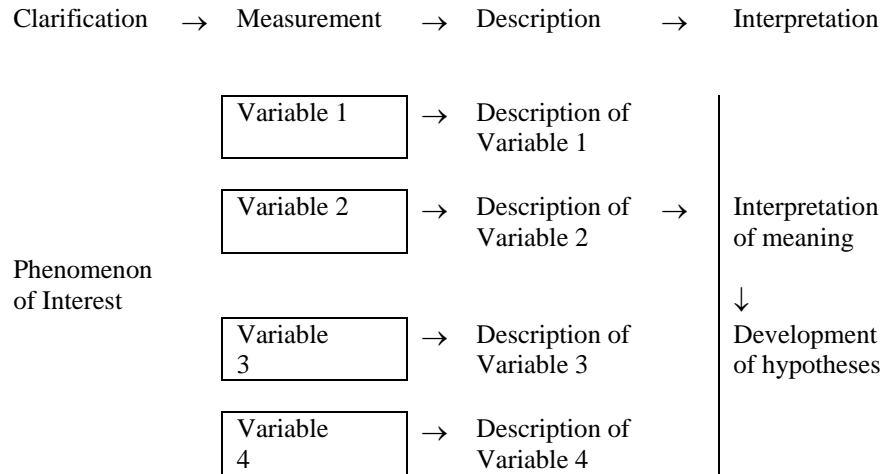
- i. Specify the sources and types of information relevant to the research problem.
- ii. Specify which approach will be used for gathering and analyzing the data.
- iii. Include the time and cost budgets, since most studies are done under these two constraints.

Types of design: Four common types of quantitative designs are used in nursing: descriptive, co relational, quasi-experimental & experimental and case study.

Descriptive studies are designed to gain more information about variables within a particular field of study. Their purpose is to provide a picture of situations as they naturally happen. No manipulation of variables is involved. **Descriptive designs** vary in level of complexity; some contain only two variables, whereas others include multiple variables.

2.3. Typical Descriptive Design

The most commonly used design in the category of descriptive studies is presented in the following Figure:



2.4. Typical Descriptive Design

The design is used to examine characteristics of a single sample. The design includes identifying a phenomenon of interest, identifying the variables within the phenomenon, developing conceptual and operational definitions of the variables, and describing the variables. The description of the variables leads to an interpretation of the theoretical meaning of the findings and development of hypotheses.

Co-relational studies examine relationships between variables. The researcher may seek to describe a relationship, predict relationships among variables, or test the relationships proposed by a theoretical proposition.

The purpose of quasi-experimental and experimental designs is to examine causality. The power of the design to accomplish this purpose is dependent on the degree to which the actual effects of the experimental treatment (the independent variable) can be detected by measurement of the dependent variable. Threats to validity are controlled through selection of subjects, manipulation of the treatment, and reliable measurement of variables. **Experimental designs**, with their strict control of variance, are the most powerful method of examining causality. **Quasi-experimental designs** were developed to provide alternative means for examining causality in situations not conducive to experimental controls. The three essential elements of experimental research are (1) random sampling; (2) researcher-controlled manipulation of the independent variable; and (3) researcher control of the experimental situation, including a control or comparison group.

2.5. Case Study Design

The case study design involves an intensive exploration of a single unit of study, such as a person, family, group, community, or institution, or a very small number of subjects who are examined intensively. Although the number of subjects tends to be small, the number of variables involved is usually large. In fact, it is important to examine all variables that might have an impact on the situation being studied.

Case studies were a commonly used design in nursing 30 years ago but appear in the literature less frequently today. Well-designed case studies are used to validate theories. Information from a variety of sources can be collected on each concept of interest using different data collection methods. This strategy can greatly expand the understanding of the phenomenon under study. Case studies are also useful in demonstrating the effectiveness of specific therapeutic techniques. In fact, the reporting of a case study can be the vehicle by which the techniques are introduced to other practitioners. The case study design also has potential for revealing important findings that can generate new hypotheses for testing. Thus, the case study can lead to the design of large sample studies to examine factors identified by the case study.

The case study design is dependent on the circumstance of the case but usually includes an element of time. The subject's history and previous behavior patterns are usually explored in detail. As the case study proceeds, the researcher may become aware of components important to the phenomenon being examined that were not originally built into the study. Both quantitative and qualitative elements are likely to be incorporated into the case study design.

Selecting a design requires an understanding of certain concepts: causality, multicausality, probability, bias, control, manipulation, and validity.

In causality, there is an assumption that situations have causes and that causes lead to effects.

Multicausality is the recognition that a number of interrelating variables can be involved in causing a particular effect.

Probability deals with the likelihood that a specific effect will occur following a particular cause.

Bias is of great concern in research because of the potential effect on the meaning of the study findings. Any component of the study that deviates

or causes a deviation from the true measurement of variables leads to distort findings.

Manipulation is moving around or controlling the movement, such as manipulating the independent variables. If the freedom to manipulate a variable is under someone else's control a bias is introduced into the study.

Control means having the power to direct manipulate factors to achieve a desired outcome. The greater the control the researcher has over the study situation, the more credible the study findings will be.

Validity is a measure of the truth or accuracy of a claim. When conducting a study, the researcher is confronted with major decisions regarding four types of validity: statistical conclusion validity, internal validity, constructs validity, and external validity. **Statistical conclusion validity** is concerned with whether the conclusions about relationships drawn from statistical analysis are an accurate reflection of real world. **Internal validity** is the extent to which the effects detected in the study are a true reflection of reality rather than a result of the effects of extraneous variables. **Construct validity** examines the fit between the conceptual and operational definitions of variables. **External validity** is concerned with the extent to which study findings can be generalized beyond the sample used in the study.

2.6. Outcome Research

In critiquing the study, it is important to identify variables not included in the design (extraneous variables) that could explain some of the variance in measurement of the study variables. In a good design, the effect of these variables on variance is controlled. Outcome research was developed to examine the end results patient care. Those promoting outcomes research demand that providers justify the selection of patient care interventions and systems of care based on evidence of improved patient lives and increased cost effectiveness.

2.7. Tools and Methods of Data Collection

Measurement Strategies in Nursing

Some of the most common measurement approaches used in nursing research, include physiologic measurement, observational measurement, interviews, questionnaires, and scales.

Empirical Phase

Physiologic Measurement

There are a variety of approaches to obtain physiologic measures. Some are relatively easy and are an extension of the measurement methods used in nursing practice, such as obtaining weight and blood pressure. Others are not difficult but require imaginative approaches to measuring phenomena that are traditionally observed in clinical practice but not measured. Some physiologic measures are obtained using self-report or paper-and-pencil scales.

Measurement of physiologic variables can be either direct or indirect. Direct measures are more valid. The measurement of arterial pressure waveforms through an arterial catheter provides a direct measure of blood pressure, whereas use of a stethoscope and sphygmomanometer provides an indirect measure

Observational Measurement

Although observational measurement is most commonly used in qualitative research, it is used to some extent in all types of studies.

Unstructured observations involve spontaneously observing and recording what is seen with minimum planning. Although unstructured observations give the observer freedom, there is a risk of loss of objectivity and a possibility that the observer may not remember all the details of the observed event.

In structured observational measurement, the researcher carefully defines what is to be observed and how the observations are to be made, recorded and coded. In most cases, a system is developed for organizing and shorting the behaviors or events being observed. Checklists are often used to indicate whether a behavior occurred or did not occur. Rating scales allow the observer to rate the behavior or event. This provides more information for analysis than dichotomous data, which indicate only that the behavior either occurred or did not occur. Observation tends to be more subjective than other types of measurement and thus is often a way to obtain important data for nursing body of knowledge.

2.8. Scales

The scale, a form of self-report, is a more precise means of measuring phenomena than the questionnaire. Most scales measure psychosocial variables. However, scaling techniques can be used to obtain self-reports on physiologic variables such as pain nausea, or functional capacity. The various items on most scales are summed to obtain a single score. These are referred to as “summated scales.” Less random and systematic error

exists when the total score of a scale is used. The various items in a scale increase the dimensions of the concept that are reflected in the instrument. The types of scales described below include **rating scales, the Likert scale, the semantic differential scale, and visual analogue scale.**

Rating scales are the crudest form of measure using scaling techniques. A rating scale lists an ordered series of categories of a variable and is assumed to be based on an underlying continuum. A numerical value is assigned to each category. The fineness of the distinctions among categories varies with the scale. The general public commonly uses rating scales. In conversations, one can hear statements such as, on a scale of one to ten, I would rank that ... This type of scale is often used in observational measurement to guide data collection.

2.9. Likert Scale

The Likert scale, which was designed to determine the opinion on or attitude toward a subject, contains a number of declarative statements with a scale after each statement. The Likert scale is the most commonly used scaling technique. The original version of the scale consisted of five categories. However, sometimes seven options are given, sometimes only four. Values are placed on each response, with a value of 1 on the most negative response and a value of 5 on the most commonly address agreement, evaluation, or frequency. Agreement response may include options such as strongly agree, agree, uncertain, disagree, and strongly disagree. Evaluation responses ask the respondent for an evaluative rating along a good/bad dimension, such as positive to negative or excellent to terrible. Frequency responses may include options such as rarely, seldom, sometimes, occasionally, and usually. The values from each item are summed to provide a total score.

Semantic Differentials

The semantic differential measures attitudes or beliefs. A semantic differential scale consists of two opposite adjectives with a 7-points scale between them. The subject is to select one point on the scale that best describes his or her view of the concept being the most negative response and 7 the most positive. The placement of negative responses to the left or right of the scale should be randomly varied to avoid global responses. The values for the scales are summed to obtain one score for each subject.

Visual Analogue Scales

The visual analogue scale is a line 100 mm long with right angle stops at either end. The line may be horizontal or vertical. Bipolar anchors are placed beyond either end of the line. This end anchor should include the

Empirical Phase

entire range of sensations possible in the phenomenon being measured (e.g., all and none, best and worst, no pain as bad as it could be possible).

The subject is asked to place a mark through the line to indicate the intensity of the stimulus. A ruler is then used to measure the distance between the left end of the line and the subject's mark. This measure is value of the stimulus. The visual analogue scales has been used to measure pain, mood, anxiety, alertness, craving for cigarettes, quality of sleep, attitudes toward environmental conditions, functional abilities, and severity of clinical symptoms.

Strategies commonly used to evaluate the reliability of scales are not useful for visual analogue scales. Because these scales are used to measure phenomena that are erratic over time, test-retest reliability is inappropriate, and because the scale consists of a single item, other methods of determining reliability cannot be used.

Interviews

An interview involves verbal communication between the researcher and the subject during which information is provided to the researcher. Although this measurement strategy is most commonly used in qualitative and descriptive studies, it can also be used in other types of studies. A variety of approaches can be used to conduct an interview, ranging from a totally unstructured interview, in which the content is similar to a questionnaire, with the possible responses to questions carefully designed by the researcher.

Unstructured interviews may be initiated by asking a broad question, such as "Describe for me your experience with" After the interview has begun, the role of the interviewer is to encourage the subject to continue talking, using techniques such as nodding the head or making sounds that indicate interest. In some case, the subject may be encouraged to elaborate further on a particular dimension of the topic of discussion.

Structured interviews involve strategies that provide increasing amounts of control by the researcher over the content of the interview. The researcher before the initiation of data collection designs questions the interviewer asks, and the order of the questions is specified. In some cases, the interviewer is allowed to explain further the meaning of the question or to modify the way in which the question is asked so that the subject can understand it better. In more structured interviews, the interviewer is required to ask the question precisely as it has been designed.

Interviewing is a flexible technique that allows the researcher to explore meaning in greater depth that can be obtained with other techniques.

Interpersonal skills can be used to facilitate cooperation and elicit more information. There is a higher response rate to interviews than to questionnaires, leading to more representative sample. Interviewer allows collection of data from subjects who are unable or unlike to complete questionnaires, such as those who are very ill or whose ability to read, write and express, themselves is marginal.

Interviews are a form of self-report, and it must be assumed that the information provided is accurate. Because of time and costs, sample size is usually limited. Subject bias is always a threat to the validity of the findings, as is inconsistency in data collection from one subject to another.

2.10. Questionnaires

A questionnaire is a printed self-reported form designed to elicit information that can be obtained through written or verbal responses of the subject. The information obtained from questionnaires is similar to that obtained by an interview, but the questions tend to have less depth. The subject is unable to elaborate on responses or ask for clarification or questions, and the data collector cannot use probing strategies. However, questions are presented in a consistent manner, and there is less opportunity for bias than in the interview. Questionnaires tend to be used in descriptive studies designed to gather a broad spectrum of information from subjects, such as facts about the subject or about persons known by the subject; opinions, levels of knowledge, or intentions of the subject. Like interviews, questionnaires can have varying degrees of structure. Some questionnaires ask **open-ended questions**, which require written responses from the subject. Others ask **closed-ended questions**, which have options selected by the researcher.

Although questionnaires can be distributed to very large samples, either directly or through the mail, the response rate to questionnaires is generally lower than that of other forms of self-report, particularly if the questionnaires are mailed. If the response rate is lower than 50%, the representativeness of the sample is seriously in question. The response rate of mailed questionnaires is usually small (25% to 30%), so the researcher is frequently unable to obtain a representative sample, even with random sampling methods. Respondents commonly fail to mark responses to all the questions, especially on long questionnaires. This can threaten the validity of the instrument.

Example of a Questionnaire

1	Do you currently smoke cigarette? a) No, B) Yes
2	How old were you when you started a) Under 15 years b) 15 years c) 16 years d) 17 years e) 18 years f) 19 years g) 20 years h) 21 years i) 22 years j) Over years
3	Before entering your basic (GENERIC) nursing education program, on the average, about how many cigarettes a day did you smoke? a) Didn't smoke at all b) Didn't smoke every day c) Less than 15 cigarettes /day d) 15-24 cigarettes/day e) 25-39 cigarettes/day f) 40 or more cigarettes/day
4	During your basic nursing (GENERIC) education program, on the average, about how many cigarettes a day did you smoke? a) Didn't smoke at all b) Didn't smoke every day c) Less than 15 cigarettes /day d) 15-24 cigarettes/day e) 25-39 cigarettes/day f) 40 or more cigarettes/day
5	How many organized programs have you attend to help you quit smoking? a) None b) One c) Two d) Three e) Four f) Five g) Six h) Seven i) More than seven
6	What is the longest single period you have stopped smoking? a) Have never stopped b) Less than a day c) Less than a week d) Less than a month e) More than 1 month but less than 1 year f) More than 1 year but less than 3 years g) 3 years or more
7.	Aside from what you think you actually could do, which would you most like to do? a) Quit smoking b) Cut down c) Cut down just a little d) Not sure at this time e) Smoke as much as now

In most questionnaires, researchers analyze individual items rather than sum the items and obtain a total score for use in data analysis. Responses to items are usually measured at the nominal or ordinal level. Because individual items may address a variety of topics associated with the research area, attempting to determine reliability-using tests of homogeneity may not be logical.

2.11. The Process of Data Collection

Data collection is the process of acquiring subjects and collecting the data needed for the study. The actual steps of collecting the data are specific to each study and are dependent on the research design and measurement techniques.

In both quantitative and qualitative research, the investigator performs five tasks during the data collection process. These tasks are interrelated and

occur concurrently rather than in sequence. The tasks include selecting subjects, collecting data in a consistent way, maintaining research controls as indicated in the study design, protecting the integrity (or validity) of the study, and solving problems that threaten to disrupt the study.

The researcher should describe the data collection process in the published study. The strategies used to approach potential subjects who meet the sampling criteria need to be clear. The number and characteristics of subjects who decline to participate in the study should be reported. The approach used to perform measurements, the timing, and the settings of measurements need to be described. The result should be a step-by-step description of exactly how, where, and in what sequence the data were collected.

In many studies, data collection forms are used to gather data. The form itself is not a measurement tool. In many cases, each item on these forms is a separate measurement. Thus, the researcher should report the source of information and describe the method and level of measurement of each item on the form. An example of a data collection form is shown below.

DATA COLLECTION FORM				
Subject Identification Number		Date		
A.	Age	B. Gender: 1. Male		2. Female
C.	Weight Kg	D. Height		cm
E.	Surgical diagnosis			
F.	Surgery Date		and Time	
G.	Narcotics Ordered after Surgery			
H.	Narcotic Administration:			
	Data	Time	Type of Narcotic	Dose
1.				
2.				
3.				
4.				
5.				
I.	Patient Instructed on Pain Scale:		Date	Time
	Comments:			
J.	Type of Treatment: 1.TENS		2. Placebo-TENS	3. No-treatment Control
K.	Treatment Implemented: DATE		Time	
	Comments:			
L.	Dressing Change: DATE		Time	
	Hours since Surgery			
M.	Score on Visual Analogue Pain Scale			
	Date		Time	
	Data Collector's Name			
			Time	

2.12. Recruiting Subjects

Subjects may be recruited only at the initiation of data collection or throughout the data collection period. The design of the study determines the methods of selecting subjects. Recruiting the number of subjects originally planned is critical because data analysis and interpretation of findings depend on an adequate sample size. Factors related to subject recruitment and selection need to be continually examined to determine possible biases in the sample obtained.

In recruiting subjects, researchers have found that direct contact with potential subjects is the most effective method, telephone contact is less effective, and mail contact is least effective. Direct contact in small groups is usually more effective for subject recruitment than is contact in large groups. The researcher must determine the most effective recruitment approach based on the purpose of the study, the type and number of subjects required, and the design of the study.

Maintaining Consistency

The key to accurate data collection in any study is consistency. Consistency involves maintaining the data collection pattern for each collection event as it was developed in the research plan. A good plan will facilitate consistency and maintain the validity of the study. However, developing a consistent plan is easier than implementing it. Deviations, though minor, need to be noted and evaluated for their impact on the interpretation of the findings. When data collectors are used in a study, they need to be trained to note deviations during the data collection process.

Maintaining Controls

Research controls are built into the plan to minimize the influence of intervening forces on study findings. Maintenance of these controls is essential; these controls are not natural in a field setting, and letting them slip is easy. In some cases, the controls slip without the researcher's realizing it. In addition to maintaining the controls identified in the plan, the researcher needs to watch continually for previously unidentified extraneous variables that might have an impact on the data being collected. These variables are often specific to a study and tend to become apparent during the data collection period. The extraneous variables identified during data collection must be considered during data analysis and interpretation. These variables also need to be noted in the research report to allow future researchers to control them.

Protecting Study Integrity

Maintaining consistency and controls during subject selection and data collection protects the integrity or validity of the study. In addition, the integrity of the study needs to be considered in a broad context. To accomplish this, we must shift from examining the elements of data collection to viewing the process of data collection as whole. Changes in one small component of data collection can modify other elements and thus alter the entire process in ways that threaten the validity of the outcomes.

Problem Solving

Problems can be perceived either as a frustration or as challenge. The fact that the problem occurred is not as important as the success of problem resolution. Therefore, the final and perhaps most important task of the data collection period may be problem resolution. Little has been written about the problems encountered by nurse researchers. The research reports often read as though everything went smoothly. The implication is that if you are a good researcher, you will have no problems, which is not true. Research journals generally do not provide sufficient space to allow description of the problems encountered, and this gives a false impression to the inexperienced researcher. A more realistic picture can be obtained through personal discussions with researchers about the data collection process.

Serendipity

Serendipity is the accidental discovery of something useful or valuable. During the data collection phase of studies, researchers often become aware of elements or relationships not previously identified. Therefore, in some published studies, the researcher has gathered data, made observations, or recorded events not originally planned. These aspects may be closely related to the study being conducted or have little connection with it. They come from the increased awareness of close observation. Because the researcher is focused on close observation, other elements in the situation can come into clear focus and take on new meaning. Serendipitous findings are important to the development of new insights in nursing. They can lead to new areas of research that generate new knowledge.



2.13. Exercise

2.13.1. Multiple-choice questions

1. Common quantitative designs are of
 - a. Two types
 - b. Three types
 - c. Four Types
 - d. Five types.
2. A blueprint for conducting a study is
 - a. A research design
 - b. Environmental design
 - c. Time design
 - d. Space design.
3. Time and cost budgets are included in
 - a. Research designs
 - b. Hypothesis
 - c. Sampling design
 - d. Experimental design.
4. Case study design involves intensive study of
 - a. A single unit
 - b. A single variable
 - c. Many units
 - d. None.
5. To get data from those whose ability to read and write is marginal we use
 - a. Questionnaires
 - b. Interviews
 - c. Likert scale
 - d. Visual analogue scale.

2.13.2. Short Questions

1. What do you mean by a research design?
2. What are descriptive studies?
3. Define observational measurement.
4. Mention the characteristics of a questionnaire.

2.13.3. Analytical questions

1. Explain the process of data collection.
2. Describe the advantages of Interviews.
3. Narrate the data collection process.

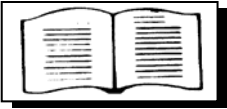
Lesson 3: Population and Sample, Methods of Sample Selection

3.1. Learning Objectives



On completion of this lesson you will be able to-

- sampling and it's necessity
- sampling theory and the related concepts
- probability and non-probability sampling methods
- the elements that influence the decision on sample size.



3.2. What is Sampling?

The process of inferring something about a large group of elements by studying only a part of it, is referred to as sampling. Most of us use sampling in our daily life, e.g. when we go to buy provisions from a grocery, we might sample a few grains of rice or wheat to infer the quality of whole bag of it.

Sampling is not the only process available for making inferences about a population. For small populations, it may be feasible and practical, and sometimes desirable to examine every member of the population e.g. for inspection of some aircraft components. This process is referred to as **census or complete enumeration** of the population.

3.3. Why Sampling?

Time taken for the Study

Inferring from a sample can be much faster than from a complete enumeration of the population because fewer elements are being studied.

Cost involved for the Study

Sampling helps in substantial cost reductions as compared to censuses.

3.4. Physical Impossibility of Complete Enumeration

In many situations the element being studied gets destroyed while being tested. The fluorescent tubes which are chosen for testing their lives, get destroyed while being tested. In such cases, complete enumeration is impossible as there would be no population left after such an enumeration.

3.5. Enough Reliability of Inferences based on Sampling

In many cases, sampling provides adequate information so that not much additional reliability can be gained with complete enumeration in spite of spending large amounts of additional money and time. It is also possible to quantify the magnitude of possible error on using some type of sampling.

3.6. Quality of Data Collected

For large populations, complete enumeration also suffers from the possibility of spurious or unreliable data collected by the enumerators. On the other hand, there is greater confidence on the purity of the data collected in sampling as there can be better interviewing, better training and supervision of enumerators, better analysis of missing data and so on.

3.7. Sampling Theory

Sampling involves selecting a group of people, events, behaviors, or other elements with which to conduct a study. A sampling plan defines the process of making the selections; the sample defines the selected group of people (or elements). Samples are expected to represent a population of people. The population might be all diabetics, all patients who have abdominal surgery, or all individuals who receive care from a nurse practitioner. However, in most cases, it would be impossible for a researcher to study an entire population for time, money and expert constraints.

3.8. Elements and Populations

An individual unit of population is called an element. An element can be a person, event, behavior, or any other single unit of a study. The sample represents a population. The population, sometimes referred to as the target population, is the entire set of individuals (or elements) who (that) met the sampling criteria.

3.9. Sampling Criteria

Sampling criteria are the characteristics essential for inclusion in the target population. These criteria might include age limitations (over 18 years), ability to speak and read English at the sixth-grade level, and not having had previous surgery etc. The sample is selected from the accessible population that meets these sampling criteria. When the study is completed, the findings are generalized to the target population that meets these sampling criteria.

3.10. Representativeness

Representativeness means that the sample, the accessible population, and the target population are alike in as many ways as possible. Representativeness to be evaluated in terms of the setting, characteristics of the subjects, and distribution of values on variables measured in the study.

A sample needs to be representative in terms of characteristics such as age, gender, ethnicity, income, and education, which often influence study variables. It is especially important that the sample be representative in relation to the variables being examined in the study.

Measurement values should also be representative. Measurement values in a study are expected to vary randomly among subjects. Random variation is the expected difference in values that occurs when different subjects from the same sample are examined. As sample size increases, random variation decreases, improving representativeness.

Systematic variation, or systematic bias—a serious concern in sampling—is a consequence of selecting subjects whose measurement values differ in some specific way from those of the population. This difference is usually expressed as difference in the average (or mean) values between the sample and the population.

The probability of systematic variation increases when the sampling process is not random. However, even in a random sample, systematic variation can occur as a consequence of potential subjects' declining participation. The greater the number who decline, the greater the possibility of a systematic bias.

3.11. Random Sampling

From a sampling theory point of view, each individual in the population should have an opportunity to be selected for the sample. One method of providing this opportunity is referred to as random sampling. The purpose of random sampling is to increase the extent to which the sample is representative of the target population.

The use of the term control group is limited to those studies using random sampling methods. If nonrandom methods are used for sample selection, the group not receiving a treatment is referred to as a comparison group because there is an increased possibility of pre-existing differences in the experimental and comparison groups.

Sampling Frames

A list of every member of the population must be acquired, using the sampling criteria to define membership. This list is referred to as the sampling frame.

Sampling Designs

A sampling design outlines strategies used to obtain a sample for a study. The design increases representativeness and decreases systematic bias.

3.12. Different Types of Sampling Designs

There are different types of sample designs based on two factors viz., the representation basis and the element selection basis. **On the representation basis**, the sample may be probability sampling or it may be non-probability sampling. Probability sampling is based on the concept of random selection, whereas non-probability sampling is ‘non-random’ sampling. **On elements selections basis**, the sample may be either unrestricted or restricted. When each sample is drawn individually from the population at large, then the sample so drawn is known as ‘unrestricted sample’. The following chart exhibits the sample designs as explained above.

Elements selection basis	Representation basic	
	Probability sampling	Non-probability sampling
Unrestricted sampling	Simple random sampling	Haphazard sampling or convenience sampling
Restricted sampling	Complex random sampling (such as cluster sampling, systematic sampling, stratified sampling etc.)	Purposive sampling (such as quota sampling, judgment sampling)

Probability Sampling Methods

Probability sampling methods have been developed to increase the representativeness of the sample. In probability sampling, every member (elements) of the population has a probability higher than zero of being selected for the sample. To achieve this probability, the sample is obtained randomly.

Simple Random Sampling

Simple random sampling is the most basic of the probability sampling plans. To achieve simple random sampling, elements are selected at random from the sampling frame.

Empirical Phase

A random sample is used when the intention is to give each item in the 'population' as much chance of being selected in the sample as every other item. A common way to conduct random sampling is to list all the 'members' or items in the population and then to use random number tables (merely a collection of random digits) or computer generated random numbers to help select the sample. Such procedures avoid various types of 'bias', which might creep into the selection of items to be included in the sample when undertaken by humans. Even the selection of numbers by individuals might be influenced by unconscious preferences for odd, even numbers, etc.

The key objective of random sampling is to obtain a sample that fairly reflects the population as a whole. Random sampling is more likely to be used when the population itself is relatively homogeneous, i.e. is composed of items of broadly the same type. Of course there may be considerable practical difficulties in actually locating and interviewing each 'member' of the population that has been randomly selected. For example, if the chosen adult is not at home or otherwise available a researcher may have to return or seek to contact that individual on another occasion. Another member of the household, etc. will not be acceptable for interviewing.

The most common method of random selections is to use a table of random numbers. To use a table of random numbers, the researcher places a pencil or finger on the table with the eyes closed. That number is the starting place. Then, moving the pencil or finger up, down, right, or left, numbers are used in order until the desired sample size is obtained. A section from a random number table is presented below:

Section from a Random Numbers Table									
06	84	10	22	56	72	25	70	69	43
07	63	10	34	66	39	54	02	33	85
03	19	63	93	72	52	13	30	44	40
77	32	69	58	25	15	55	38	19	62
20	01	94	54	66	88	43	91	34	28

To select a sample of size 5 from a population of size 100, if the pencil were initially placed on 58 which is the fourth column from the left and the fourth row down; if a decision were made to go across the columns to the right, the subject numbers would be 58, 25, 15, 55, and 38. However, computer programs are also available for selecting samples from populations.

The main characteristics of simple random samples are:

Random Sampling

- i. Each item selected has an equal chance of being drawn
- ii. Usually adopted when the population is largely homogeneous, i.e. when it is difficult to distinguish between items
- iii. Implementation often involves the use of computer-generated random numbers
- iv. Selection is unbiased
- v. The chosen items need to be located, then questioned or measured
- vi. A major drawback is that a population listing is required.

3.13. Stratified Random Sampling

Stratified random sampling is used in situations in which the researcher knows some of the variables in the population that are critical to achieving representativeness. Variables commonly used for stratification include age, gender, ethnicity, socioeconomic status, diagnosis, geographic region, type of institution, type of care, and site of care. Stratification ensures that all levels of the identified variables will be adequately represented in the sample. With stratification, the researcher can use a smaller sample size and achieve the same degree of representativeness in relation to the stratified variable as a large sample acquired through simple random sampling.

The main characteristics of stratified random sampling are:

- i. Used when the population has a number of identifiable attributes
- ii. Populations stratified in this way are known as heterogeneous
- iii. The composition of the sample must reflect the attribute present in the population, e.g. the proportion of low, middle-and high-income earners
- iv. Individuals or items within each stratum may still be selected randomly
- v. A stratified sample is free from selective bias, since it reflects the proportions of any given attribute present in the population as a whole.

3.14. Proportional Stratified Sampling

If the different strata have similar variances of the characteristic being measured, then the statistical efficiency will be the highest if the sample sizes for different strata are in the same proportion as the size of the respective stratum in the population. Such a design is called proportional stratified sampling.

Empirical Phase

In this method after defining the strata, a simple random sample is picked up from each of the strata. If we want to have a total sample of size 100, this number is allocated to the different strata in proportion to the size of the stratum in the population.

If, we want to pick up a proportional stratified sample of size n from a population of size N , which has been stratified to p different strata with sizes $N_1, N_2, N_3, \dots, N_p$, respectively, then the sample sizes for different strata, viz $N_1, N_2, N_3, \dots, N_p$, will be given by

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \dots = \frac{n_p}{N_p} = \frac{n}{N}$$

Calculation of **Proportional Stratified Sample** sizes are given below:

Stratum No (i)	No. of Elements in stratum (N_i)	Sample size (n_i)	Sampling Ratio (n_i/N_i)
1	200	10	1/20
2	300	15	1/20
3	<u>500</u>	<u>25</u>	<u>1/20</u>
Total	1000	50	1/20

3.15. Disproportional Stratified Sampling

If the different strata in the population have unequal variances of the characteristic being measured, then the sample size allocation decision should consider the variance as well. It would be logical to have a smaller sample from a stratum where the variance is smaller than from another stratum where the variance is higher. In fact, if $\sigma_1^2, \sigma_2^2, \sigma_3^2, \dots, \sigma_p^2$ are the variance of the p strata respectively, then the statistical efficiency is the highest when

$$\frac{n_1}{N_1\sigma_1} = \frac{n_2}{N_2\sigma_2} = \dots = \frac{n_p}{N_p\sigma_p}$$

Where the other symbols have the same meaning as in the previous example.

Calculation of **Disproportional Stratified Sample** sizes are given below:

Stratum No (i)	No. of Elements in stratum (N_i)	Stratum Variance σ_i^2	Stratum s.d. (σ_i)	Sample size (n_i)	Sampling Ratio (n_i/N_i)
1	200	2.25	1.5	13	0.065
2	300	4.00	2.0	26	0.087
3	<u>500</u>	0.25	0.5	<u>11</u>	0.022
Total	1000			50	

3.16. Cluster Sampling

In cluster sampling, a sampling frame is developed that includes a list of all the states, cities, institutions, or organizations with which elements of the identified population would be linked. A random sample of these states, cities, institutions, or organization would then be used in the study. In some case, this randomized selection continues through several stages and is then referred to as “multistage sampling.”

Cluster sampling is used in two situations. The first is one in which the researcher considers if necessary to obtain a geographically dispersed sample but realizes that a simple random sample would be prohibitive in terms of travel time and cost. The second situation is one in which individual elements of a sampling frame are not available.

The main characteristics of cluster sampling are:

- i. Items chosen in cluster rather than individually. Example, cluster might be all residents in a particular road or group of roads.
- ii. Useful methods where the population is widely spread geographically but where the various clusters are broadly representative of that population.

Multi-stage Sampling

- i. Usually involves sampling of subjects with a given attribute
- ii. Often occurs where there is a wide geographical spread of such subjects which makes sampling expensive
- iii. Whole geographical area is therefore divided into regions
- iv. A small number of such regions is then selected randomly
- v. Region selected are further broken down into sub-regions from which a random sample is selected
- vi. Sub-regions selected are further broken down into unit (e.g. towns or streets), from which a random sample is again selected
- vii. Eventually individual households or persons with a given attribute are identified in specific towns or streets
- viii. Costs of interviewing are then much reduced, though some risks of bias in procedure especially if only small numbers of regions, sub-regions, etc. are selected.

Systematic Sampling

Systematic sampling can be conducted when an ordered list of all members of the population is available. The process involves selecting every k th individual on the list, using a starting point selected randomly. If the initial starting point is not random, the sample is not a probability sample. To use this design, the researcher must know the number of elements in the population and the size of the sample desired. The population size is divided by the desired sample size, giving k , the size of the gap between elements selected from the list. For example, if the population size was $N = 1,242$ and the desired sample size was $n = 50$, then $k = 24$. Every 24th person on the list would be included in the sample.

3.17. Nonprobability Sampling Methods

In nonprobability sampling, not every element of the population has an opportunity for selection in the sample. Although this approach increases the possibility of samples that are not representative, it has been commonly used in nursing studies.

The four nonprobability designs discussed here are convenience sampling, quota sampling, purposive sampling, and network sampling.

Convenience Sampling

Convenience sampling (also called “accidental sample”) is considered a poor approach because it provides little opportunity to control for biases; subjects are included in the study merely because they happened to be in the right place at the right time. Available subjects are simply entered into the study until the desired sample size is reached. Convenience samples are inexpensive, accessible, and usually require less time to acquire than other types of samples. They provide means to conduct studies on topics that cannot be examined with probability sampling. However, this type of sampling should be used only when it is impossible to obtain a sample by other means. Most quasi-experimental studies and clinical trials in both medicine and nursing use convenience sampling.

Quota Sampling

Quota sampling uses a convenience sampling technique with an added features strategy to ensure the inclusion of subject types who are likely to be underrepresented in the convenience sample, such as females, minority groups, the aged, the poor, the rich, and the undereducated. The goal of quota sampling is to represent the proportions of subgroups present in the population.

The use of quota sampling is widespread in market research. Here the intention is often to deliberately introduce selective bias into the samples, in the sense that attributes of the members or items selected will represent the choice of the samples rather than the attributes of the population as a whole. In this sense there is no attempt to seek a representative or unbiased sample from the population. There is usually no attempt to use random sampling within the quota selected, as often happens within the different strata in stratified sampling. It is often left to the discretion of interviewer, etc. to include specified numbers (quotas) of subjects possessing given attributes within the sample. To do so accurately may be costly, requiring highly trained interviewers.

The main characteristics of quota sampling are:

- i. Widely used in market research
- ii. Sample includes a specified number of quota of subjects with given attributes
- iii. Interviewers must be highly trained as they are often responsible for identification and selection of respondents
- iv. A 'biased sample therefore results, but one which may be useful in representing the customer seen as most likely to purchase the firms products.

Purposive Sampling

Purposive sampling, sometimes referred to as “judgmental or theoretical sampling,” involves the conscious selection by the researcher or certain subjects or elements to include in the study. Efforts might be made to include typical subjects or situations.

Network Sampling

Network sampling, sometimes referred to as “snowballing,” holds promise for locating samples difficult or impossible to obtain in other ways. Network sampling takes advantages of social networks and the fact that friends tend to have characteristics in common. When the researcher has found a few subjects with the necessary criteria, he or she asks their assistance in getting in touch with others with similar characteristics. This strategy is particularly useful for finding subjects in socially devalued populations such as alcoholics, child abusers, sex offenders, drug addicts, and criminals. These individuals are seldom willing to make them selves known.

Empirical Phase

Sample Size

One of the most troublesome questions that arise during the critique of a study is whether the sample size was adequate. If the study was designed to make comparisons and significance was found, the sample size was adequate. Questions about the adequacy of the sample size occur only when no significance was found.

Currently, the adequacy of the sample size is evaluated using a power analysis. Power is the capacity of the study to detect differences or relationships that actually exist in the population. Expressed another way, it is the capacity to correctly reject a null hypothesis. The minimum acceptable level of power for a study is .80. This power level results in a 20% chance of a Type II error, in which the study fails to detect existing effects (differences or relationships).

Other factors that influence the adequacy of sample size include effect size, type of study, number of variables, sensitivity of the measurement tools, and data analysis techniques.

Effect Size

The effect is the presence of the phenomenon examined in a study. Effect size is the extent to which the null hypothesis is false. In a study in which two populations are compared, the null hypothesis states that the difference between the two populations is zero. However, if the null hypothesis were false, there would be an effect. If the null hypothesis is false, it is false to some degree; this is the effect size. The statistical test tells you that there is a difference between groups. The effect size tells you how much difference there is between groups.

When the effect size is large (e.g., considerable difference between groups), detecting it is easy and requires only a small sample; when the effect size is small, detecting it is more difficult and requires larger samples. Broadly, a small effect size would be about .2, a medium effect size .5, and a large effect size .8. Effect size is smaller with a small sample, and thus effects are more difficult to detect. Increasing the sample size also increases the effect size, making it more likely that the effect will be detected.

Types of Study

Case studies tend to have very small samples. Comparisons between groups are not performed, and problems related to sampling error and generalization has little relevance for these studies. A small sample size may be more useful in examining the situation in depth from various

perspectives. Descriptive studies and correlational studies often require very large samples.

Number of Variables

As the number of variables under study increases, the sample size needed may increase. The inclusion of multiple dependent variables also increases the sample size.

Measurement Sensitivity

Well-developed instruments measure phenomena with precision. A thermometer, for example, measures body temperature precisely. Tools measuring psychosocial variables tend to be less precise. Variance tends to be higher in a less well-developed tool than in one that is well developed. As variance in instrument scores increases, the sample size needed to obtain significance increases.

Data Analysis Techniques

Data analysis techniques vary in their ability to detect differences in the data. Statisticians refer to this as the “power of the statistical analysis” There is also an interaction between the measurement sensitivity and the power of the data analysis technique. The power of the analysis technique increases as precision in measurement increases. Larger samples are needed when the power of the planned statistical analysis is weak.



3.18. Exercise

3.18.1. Multiple-choice questions

1. Sample is a part of
 - a. Population
 - b. Country
 - c. World
 - d. Space.

2. The purpose of random sampling is to ensure
 - a. Non zero probability for each element
 - b. Complete counting
 - c. Subjective inference
 - d. None of the above.

Empirical Phase

3. A sampling design increases
 - a. Ambiguity
 - b. Bias
 - c. Cost
 - d. Representativeness.
4. Sampling designs are divided into
 - a. Two categories
 - b. Three categories
 - c. Four categories
 - d. Five categories.
5. Within stratum elements are
 - a. Homogeneous
 - b. Heterogeneous
 - c. Continuous
 - d. Clusters.

3.18.2. Short questions

1. What is sampling?
2. Mention the necessity of sampling.
3. What are non-probability samplings?
4. Mention the factors that influence sample size.

3.18.3. Analytical questions

1. Explain different types of designs based on representation and element selection basis.
2. Narrate the process of taking a simple random sample.
3. Describe the proportional stratified sampling.

Lesson 4: Methodology and Action Plan for Data Collection

4.1. Learning Objectives



On completion of this lesson you will be able to-

- discuss the necessity and usefulness of data collection
- distinguish between primary and secondary data
- describe different methods of collecting primary data with their merits and demerits
- identify the sources of secondary data
- examine the reliability, suitability and adequacy of secondary data
- understand how to select appropriate method of data collection.



4.2. Necessity and Usefulness of Data Collection

The necessity and usefulness of information gathering or data collection cannot be overemphasized in formulating policies. For example, the government must be aware of the actual scenario of the acceptance of family planning before it can formulate any policy in this matter. The components of this scenario are provided by appropriate data to be collected from various families. In industrial disputes regarding wages, cost of living index, a data based indicator of inflations is often accepted as a guideline for arbitration. Similarly in formulating health policies we must have at our disposal related data.

In short no decision can be made in a casual manner in the highly involved environment prevailing in this age. It is through appropriate data and their analysis that the decision maker becomes equipped with proper tools of decision-making.

4.3. Method of Data Collection

While deciding about the method of data collection to be used for the study, the researcher should keep in mind that there are two types of data viz., primary and secondary. **The primary data** are those, which are collected afresh and for the first time, and happen to be original in character. **The secondary data**, on the other, are those which have already been collected by someone else and which have already been passed through the statistical process.

Collection of Primary Data

There are several methods of collecting primary data, particularly in surveys and descriptive researches. Important ones are: i) observation

Empirical Phase

method, ii) interview method, iii) through questionnaires, iv) through schedules, v) case study method, and vi) other methods which include a) warranty cards; b) distributor audits, c) pantry audits, d) consumer panels, e) using mechanical devices, f) through projective techniques, g) depth interviews, and h) content analysis. We briefly describe each method separately.

Under the observation method, the information is sought by way of investigator's own direct observation without asking the respondent. The main advantage of this method is that subjective bias is eliminated, if observation is done accurately. Secondly, the information obtained under this method relates to what is currently happening; it is not complicated by either the past behavior or future intentions or attitudes. Thirdly, this method is independent of respondents' willingness to respond and as such is relatively less demanding of active cooperation on the part of respondents as happens to be the case in the interview or questionnaire method.

However, observation method has various limitations. Firstly, it is an expensive method. Secondly, the information provided by this method is very limited. Thirdly, sometimes-unforeseen factors may interfere with the observational task. At times, the fact that some people are rarely accessible to direct observation creates obstacle for this method to collect data effectively.

Sometimes we talk of **controlled and uncontrolled observation**. If the observation takes place in the natural setting, it may be termed as uncontrolled observation, but when observation takes place according to definite pre-arranged plans, involving experimental procedure, the same is then termed controlled observation. In non-controlled observation, no attempt is made to use precision instruments. The major aim of this type of observation is to get a spontaneous picture of life and persons. It has a tendency to supply naturalness and completeness of behavior, allowing sufficient time for observing it. But in controlled observation, we use mechanical (or precision) instruments as aids to accuracy and standardization. Such observation has tendency to supply formalized data upon which generalizations can be built with some degree of assurance.

4.4. Interview Method

This method can be used through personal interviews and, if possible, through telephone interviews.

Personal interview method requires a person known as the interviewer asking questions generally in a face-to-face contact to the other person or persons.

The chief merits of the interview method are as follows:

- i. More information and that too in greater depth can be obtained.
- ii. Interviewer by his own skill can overcome the resistance, if any, of the respondents; the interview method can be made to yield an almost perfect sample of the general population.
- iii. There is greater flexibility under this method as opportunity to restructure questions is always there, specially in case of unstructured interviews.
- iv. Observation method can also be applied to recording verbal answers to various questions.
- v. Personal information can as well be applied easily under this method.
- vi. Samples can be controlled more effectively as there arises no difficulty of the missing returns: non-response generally remains very low.
- vii. The interviewer can usually control which persons (s) will answer the questions.
- viii. The interviewer may catch the information off-guard and thus may secure the most spontaneous reactions than would be case if mailed questionnaire is used.
- ix. The language of the interview can be adopted to the ability or educational level of the person interviewed and as such misinterpretations concerning questions can be avoided.
- x. The interviewer can collect supplementary information about the respondent's personal characteristics and environment, which is often of great value in interpreting results.

But there are also certain weaknesses of the interview method. Among the important weaknesses, mention may be made of the following:

- i. It is very expensive method, specially when large and widely spread geographical sample is taken.
- ii. There remains the possibility of the bias of interviewer as well as that of the respondent.
- iii. Certain types of respondents such as important officials or executives or people in high income groups may not be easily approachable under this method and to that extent the data may prove inadequate.

Empirical Phase

- iv. This method is relatively more time consuming, specially when the sample is large and recalls upon the respondents are necessary.
- v. The presence of the interviewer on the spot may over-stimulate the respondent, sometimes even to the extent that he may give imaginary information just to make the interview interesting.
- vi. Under the interview method the organization required for selecting, training and supervising the field-staff is more complex with formidable problems.
- vii. Interviewing at times may also introduce systematic errors.
- viii. Effective interview presuppose proper rapport with respondents that would facilitate free and frank responses. This is often a very difficult requirement.

4.5. Telephone Interviews

This method of collecting information consists in contacting respondents on telephone itself. It is not a very widely used method, but plays important part in industrial surveys, particularly in developed regions. The chief merits of such a system are:

- i. It is more flexible in comparison to mailing method.
- ii. Recall is easy; callbacks are simple and economical.
- iii. Interviewer can explain requirements more easily.
- iv. At times, access can be gained to respondent who otherwise cannot be contacted for one reason or the other.
- v. Representative and wider distribution of sample is possible.

But this system of collecting information is not free from demerits. Some of these may be highlighted.

- i. Surveys are restricted to respondents who have telephone facilities.
- ii. It is not suitable for intensive surveys where comprehensive answers are required to various questions.
- iii. Questions have to be short and to the point; probes are difficult to handle.

4.6. Collection of Data Through Questionnaires

A questionnaire consists of a number of questions printed or typed in a definite order on a form or set of forms. The questionnaire is mailed to respondents who are expected to read and understand the questions and write down the reply in the space meant for the purpose in the questionnaire itself. The respondents have to answer the questions on their own.

The method of collecting data by mailing the questionnaires to respondents is most extensively employed in various economic and business surveys. The main merits of this method are:

- i. There is low cost even when the universe is large and is widely spread geographically.
- ii. It is free from the bias of the interviewee; answers are in respondents' own words.
- iii. Respondents have adequate time to give well thought out answers.
- iv. Respondents, who are not easily approachable, can also be reached conveniently.
- v. Large samples can be made use of and thus the results can be made more dependable and reliable.

The main demerits of this method are:

- i. Low rate of return of the duly filled in questionnaires; bias due to no-response is often indeterminate.
- ii. It can be used only when respondents are educated and cooperating.
- iii. The control over questionnaire may be lost once it is sent.
- iv. There is inbuilt inflexibility because of the difficulty of amending the approach once questionnaires have been dispatched.
- v. There is also the possibility of ambiguous replies or omission of replies altogether to certain questions; interpretation of omissions is difficult.
- vi. It is difficult to know whether willing respondent are truly representative.
- vii. This method is likely to be the slowest of all.

Empirical Phase

Comparison the above Methods of Data Collection are presented below:

Personal Interview with Questionnaire		Advantages Mailed Questionnaire		Telephone Interview	
i)	Most flexible method of obtaining data	i)	Wide and more representative coverage possible at the same cost	i)	Wide and more representative coverage possible at the same cost
ii)	Identity of respondent known	ii)	No field Staff is required	ii)	No field Staff is required
iii)	No response very low	iii)	Cost per questionnaire relatively low	iii)	Cost of response low
iv)	Supervision and control possible	iv)	No bias of interviewers	iv)	Quick way of obtaining data
		v)	Respondents can answer question directly		
		Disadvantages			
i)	Most expensive method of collection data	i)	High degree of non response	i)	Non telephone owners and those with unlisted numbers cannot be reached
ii)	Considerable supervision necessary	ii)	Questions which require probing cannot be asked	ii)	Interview period is short
iii)	Bias of the investigators present in the responses	iii)	Slowest of all methods of data collection	iii)	Questions which require probing cannot be asked
		iv)	Assumes that the respondent is educated.		

Following table gives the Preferred Data Collections Methods under different conditions of funds and time:

Funds	Time	Type of Data	Personal Interview	Mail	Telephone
Restricted	Restricted	Few Items			×
Restricted	Restricted	Much Information	×		
Restricted	Ample	Few Items		×	×
Restricted	Ample	Much Information	×	×	
Ample	Restricted	Few Items			×
Ample	Restricted	Much Information	×		
Ample	Ample	Few Items			×
Ample	Ample	Much Information	×		

4.7. Collection of Data Through Schedules

This method of data collection is very much like the collection of data through questionnaire, with little difference, which lies in the fact that schedules (proforma containing a set of questions) are being filled in by the enumerator who is specially appointed for the purpose.

4.8. Difference Between Questionnaires and Schedules

The important points of difference are as under:

- i. The questionnaire is generally sent through mail to information to be answered as specific in covering letter, but otherwise without further assistance from the sender. The research worker or the enumerator, who can interpret questions when necessary, generally fills out the schedule.
- ii. To collect data through questionnaire is relatively cheap and economical since we have to spend money only in preparing the questionnaire and in mailing the same to respondents. To collect data through schedules is relatively more expensive since considerable amount of money has to be spent in appointing enumerator and in imparting training to them. Money is also spent in preparing schedules.
- iii. Non-response is usually high in case of questionnaire as many people do not respond and many return the questionnaire without answering all questions. Bias due to non-response often remains indeterminate. As against this, non-response is generally very low in case of schedules because enumerators who are able to get answerers to all questions fill these. But there remains the danger of interviewer bias and cheating.
- iv. In case of questionnaire, it is not always clear as to who replies, but in case of schedule the identify of respondent is known.

4.9. Case Study Method

The case study method is a very popular form of qualitative analysis and involves a careful and complete observation of a social unit, be that unit a person, a family, an institution, a cultural group or even the entire community. It is a method of study in depth rather than breadth. The case study places more emphasis on the full analysis of a limited number of events or conditions and their interrelations. The case study deals with the processes that take place and their interrelationship. Thus, case study is essentially an intensive investigation of the particular unit under consideration. The object of the case method is to locate the factors that account for the behavior-patterns of the given units as an integrated totality.

4.10. Editing of Primary Data

Editing involve reviewing the data collected by investigators to ensure maximum accuracy and unambiguity. It should be done as soon as possible after the data have been collected. If the size of the data is relatively small, it is desirable that only one person edit all the data for the entire study. The different steps of editing are indicated below.

Empirical Phase

- i. Checking legibility:** Obviously, the data must be legible to be used. If a response is not presented clearly, the concerned investigator should be asked to rewrite it.
- ii. Checking completeness:** An omitted entry on a fully structured questionnaire may mean that no attempt was made to collect data from the respondent or that the investigator simply did not record the data. If the investigator did not record the data, prompt editing and questioning of the investigator may provide the missing item. If an entry is missing because of the first possible cause, there is not much that can be done, except to make another attempt to get the missing data. Obviously, this requires knowing why the entry is missing.
- iii. Checking consistency:** The editor should examine each questionnaire to check inconsistency or inaccuracy if any, in the statement. The income and expenditure figures may be unduly inconsistent. The age and the date of birth may disagree. The area of an agriculture plot may be unduly large. The concerned investigators should be asked to make the necessary corrections. If there is any repetitive response pattern in the reports of individual investigators they may represent investigator bias or perhaps attempted dishonesty.

4.11. Collection of Secondary Data

Secondary data means that are already available i.e., they refer to the data, which have already been collected and analyzed by someone else. When the researcher utilizes secondary data, then he has to look into various sources from where he can obtain them. In this case he is certainly not confronted with the problems that are usually associated with the collection of original data. Secondary data may either be published data or unpublished data. Usually published data are available in: a) various publications of the central, state or local governments; b) various publications of foreign governments or of international bodies and their subsidiary organizations; c) technical and trade journals; d) books, magazines and newspapers; e) reports and publications of various associations connected with business and industry, banks, stock exchanges, etc; f) reports prepared by research scholars, universities, economists, etc. in different fields; and g) public records and statistics, historical documents, and other sources of published information. The sources of unpublished data are many; they may be found in diaries, letters, unpublished biographies and autobiographies and also may be available with scholars and research workers, trade associations, labor bureaus and other public/private individuals and organizations.

4.12. Scrutiny of Secondary Data

Primary data are to be scrutinized after the interviewers complete the questionnaires. Likewise, the secondary are to be scrutinized before they are compiled from the source. The scrutiny should be made to assess the **suitability, reliability, adequacy and accuracy** of the data to be compiled and to be used for the proposed study.

- i. **Suitability:** The compiler should satisfy him self that the data contained in the publication would be suitable for his study. In particular, the conformity of the definitions, units of measurement and time frame should be checked.
- ii. **Reliability:** The reliability of the secondary data can be ascertained from the collecting agency, mode of collection and the time period of collection. For instance, secondary data collected by a voluntary agency with unskilled investigators are unlikely to be reliable.
- iii. **Adequacy:** The source of data may be suitable and reliable but the data may not be adequate for the proposed enquiry. The original data may cover a bigger or narrower geographical region or the data may not cover suitable period.
- iv. **Accuracy:** The user must be satisfied about the secondary data. The process of collecting raw data, the reproduction of processed data in the publication, the degree of accuracy desired and achieved should also be satisfactory and acceptable to the researcher.

4.13. Selection of Appropriate Method for Data Collection

There are various methods of data collection. As such the researcher must judiciously select the method/methods for his own study, keeping in view the following factors:

- i. **Nature, scope and object of enquiry:** This constitutes the most important factor affecting the choice of a particular method. The method selected should be such that it suits the types of enquiry that is to be conducted by the researcher. This factor is also important in deciding whether the data already available (secondary data) are to be used or the data not yet available (primary data) are to be collected.
- ii. **Availability of funds:** Availability of funds for the research project determines to a large extent the method to be used for the collection of data. When funds at the disposal of the researcher are very limited, he will have to select a comparatively cheaper method, which may not be as efficient and effective as some other costly method. Finance, in fact, is a big constraint in practice and the researcher has to act within this limitation.

Empirical Phase

- iii. **Time factor:** Availability of time has also to be taken into account in deciding a particular method of data collection. Some methods take relatively more time, whereas with others the data can be collected in a comparatively short duration. The time at the disposal of the researcher, thus, affects the selection of the method by which the data are to be collected.
- iv. **Precision required:** Precision required is yet another important factor to be considered at the time of selecting the method of collection of data.

4.14. Coding of Data

Coding is the process of assigning some symbols (either) alphabetical or numerical or (both) to the answers so that the responses can be recorded into a limited number of classes or categories. The classes should be appropriate to the research problem being studied. They must be exhaustive and must be mutually exclusive so that the answer can be placed in one and only one cell in a given category. Further, every class must be defined in terms of only one concept.

The coding is necessary for the efficient analysis of data. The coding decisions should usually be taken at the designing stage of the questionnaire itself so that the likely responses to questions are pre-coded. This simplifies computer tabulation of the data for further analysis. It may be noted that any errors in coding should be eliminated altogether or at least be reduced to the minimum possible level.

4.15. Classification of Data

In most research studies, voluminous raw data collected through a survey need to be reduced into homogeneous groups for any meaningful analysis. This necessitates classification of data, which in simple terms is the process of arranging data in groups or classes on the basis of some characteristic. Classification helps in making comparisons and drawing meaningful conclusions. Classification can either be according to attributes or according to numerical characteristics.

Statistical Series

A series is defined as logical or systematic arrangement of observations or items. When the attributes or things are counted, measured or weighed and arranged in an orderly manner, say either descending or ascending order, they constitute a series. When the statistical data pertains to time, the series is said to be **historical or time series**. The important factor in such series is the chronology. When the data pertains to space, the series is referred to as special, and is also known as **geographical series**. When the

data refers to physical conditions such as height, weight, age, etc., the series is referred to as **condition series**.

4.16. Tables and Graphs as Data Presentation Devices

Statistical data can be presented in the form of tables and graphs. In the tabular form, the classification of data is made with reference to time or some other variables. The graphs are used as visual form of presentation of data.

The tabulation is used for summarization condensation of data. It aids in analysis of relationships, trends and other summarization of the given data. The tabulation may be simple or complex. Simple tabulation results in one-way tables, which can be used to answers questions related to one characteristic of data. The complex tabulation usually results in two way tables, which give information about two interrelated characteristics of the data; three way tables which give information about three interrelated characteristics of data; and still higher order tables, which supply information about several interrelated characteristics of data.

Following are the important characteristics of a table:

- i. Every table should have a clear and concise title to make it understandable without reference to the text. Thus title should always be just above the body of the table.
- ii. Every table should be given a distinct number to facilitate easy reference.
- iii. Every table should have captions (column headings) and stubs (row headings) and they should be clear and brief.
- iv. The units of measurements used must always be indicated.
- v. Source or sources from where the data in the table have been obtained must be indicated at the bottom of the table.
- vi. Explanatory footnotes, if any, concerning the table should be given beneath the table along with reference symbol.
- vii. The columns in the tables may be numbered to facilitate reference.
- viii. Abbreviations should be used to the minimum possible extent.
- ix. The tables should be logical, clear, accurate and as simple as possible.
- x. The arrangement of the data categories in a table may be chronological, geographical, alphabetical or according to magnitude to facilitate comparison.
- xi. The table must suit the needs and requirements of the research study.
- xii. Finally, several types of graph or charts are used to present tabulated data.



4.17. Exercise

4.17.1. Multiple-choice questions

1. Data are of
 - a. Two types
 - b. Three types
 - c. Four types
 - d. Five types.
2. Primary data are collected through
 - a. Questionnaires
 - b. Schedules
 - c. Both a and b
 - d. None.
3. Schedules are filled in by the
 - a. Respondent
 - b. Director
 - c. Enumerators
 - d. Supervisor.
4. Editing of data ensures
 - a. Accuracy
 - b. Randomness
 - c. Representativeness
 - d. Unbiasedness.

4.17.2. Short questions

1. Define different types of data.
2. What do you mean by reliability of secondary data.
3. Mention preferred data collection methods with restricted funds.
4. What is coding? Why do we need coding?

4.17.3. Analytical questions

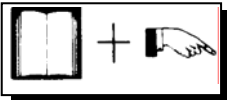
1. Mention the merits of collecting data by questionnaires.
2. Describe the characteristics of a table.
3. Give a comparative study between questionnaires and schedules.
4. Narrate the factors for selecting appropriate method for data collection.

Unit 4: Interpretive Phase

Lesson 1: Basic Statistical Concept and Common Terms

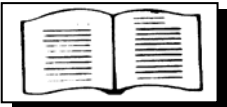
Frequency distribution, mean, median, mode and standard deviation

1.1. Learning Objectives



On completion of this lesson you will be able to-

- basic statistical concept and tools
- frequency distribution
- mean, median, mode and standard deviation.



1.2. Basic Statistical Concept and Common Terms

The word “statistics” seems to have obtained from the Latin word “status” or the Italian word “statistik” each of which means Political State”. In ancient time, the government used to collect information about total population, land, wealth, total number of employees, soldiers etc. to have the idea of the manpower of the country for formulation of administrative set-up, fiscal, new taxes, levies and military policies of the government.

Different authors defined statistics in a number of ways. Croxton and Cowden have given a very simple and concise definition of statistics. In their view, statistics may be defined as a science of collection presentation, analysis and interpretation of numerical data. This definition clearly points out four stages in a statistical investigation, namely: (i) collection of data (ii) presentation of data (iii) analysis of data and (iv) interpretation of data.

However, to the above stages, one more stage may be added and that is the organization of data. Thus, statistics may be defined as the science of collection, organization, presentation, analysis and interpretation of numerical data.

Now-a-days statistics is not only used for collecting numerical data but also to develop sound techniques for their handling, analysis and drawing valid inference from them. It is now used widely in different spheres of life social, political and also in different fields such as Agriculture, Planning, Biology, Psychology, Education, Economics, and Business Management etc.

1.3. Common Statistical Term/Tools

Variable: Anything that varies within limit such as height, weight, etc is called variable.

Frequency: Number of persons in each group is called frequency.

Normal curve: Large number of sample with small class intervals gives a frequency curve, which is symmetrical in nature, is called normal curve.

Observation: An event and its measurements such as blood pressure (event) and 120/80 mm of Hg (measurement) observation unit. The source that gives observations such as object, person etc.

Data: A set of values recorded on one or more observational units.

Population: It is an entire group of people or study element, persons, things or measurements.

Sample: It is the part of a whole group by which information is obtained on experiment on the whole group or part of while group, which is usually, represent the whole group.

Sampling: It is the process of obtaining information about the whole group by examining only a part of the group.

Variance: Estimation of variation of the sample.

Range: It is the absolute difference between the largest and smallest values in a set of data.

Parameter: It is a summary value or constant of a variable that describes the population such as mean, variance, correlation co-efficient, proportion, etc.

Percentiles: Percentiles are values in a series of observations arranged in ascending order of magnitude, which divide the distribution into 100 equal parts. 50th percentile will have 50% observations on either side. 10th percentile will have 10% observation to the left and 90% to the right e.g., population age within 3½ years in Bangladesh is 10th percentile of entire population it means 10% of the population is below 3½ years of age and 90% is above that age.

1.4. Frequency Distribution

The representation of data in a table, which describes the pattern of observations through its range, is called frequency distribution.

Construction of a Frequency Distribution

Following are the steps for the constructions of a frequency distribution.

1. Find out range by subtracting the lowest value from the highest value of the variable x.
2. The number of class intervals should not be too large or too small considering practical situation. Having fixed number of classes, divide the range by it and the nearest integer to this value is class interval. The class intervals should be exhaustive, mutually exclusive and usually of equal length.
3. The table will have 3 columns namely-class interval, tally marks and frequency. The first class interval will start with the smallest value and continue until the interval with the highest value of given series of data is reached.
4. Give tick mark to each of the values of the original data and a notation (/) which is usually called tally mark is put against each value of the variable x of the appropriate class interval. Having occurred 4 times, 5th occurrence is represented by putting a cross tally (⊥) on the first four tallies. Thus finish all the values one after another. In case of continuous frequency distribution, the variable x should follow either lower limit $\leq x <$ upper limit or lower limit $< x \leq$ upper limit. The former limit is usually considered.
5. Count the number of tally marks corresponding to each class interval and write the result in the respective frequency column.

Example- Prepare a frequency distribution table base on marks of statistics of 100 students of a certain university.

54	32	38	44	48	41	30	43	46	41
47	32	26	25	41	31	51	43	45	32
51	50	34	38	44	38	54	32	39	41
42	38	41	25	45	36	40	50	52	30
41	32	27	30	40	42	52	48	49	37
48	39	26	54	47	49	38	26	27	49
47	49	32	51	49	33	47	55	25	28
37	36	44	53	48	54	29	37	39	40
50	30	55	48	36	34	27	53	28	52
47	35	46	48	32	29	54	49	47	53

Interpretive Phase

A frequency distribution table can be prepared from above data as follows-

Class interval of marks	Tally marks	Frequency
25-29		13
30-34		15
35-39		15
40-44		16
45-49		22
50-54		17
55-59		2

1.5. Graphical Representation of Frequency Distribution

Graphical representation of a frequency distribution is more effective than tabular representation. Diagrams are essential to convey the statistical information to the general public. It also facilitates the comparison of two or more frequency distributions.

The following types of graphs are generally used to represent the frequency distribution.

1. Bar diagram
2. Histogram
3. Frequency polygon
4. Cumulative frequency polygon
5. Cumulative frequency curve or ogive
6. Pie diagram

1. Bar diagram: Diagram may be of following types.

i. **Simple bar diagram:** A simple bar diagram is used to represent only one variable. The bars are of same width and varies only the length.

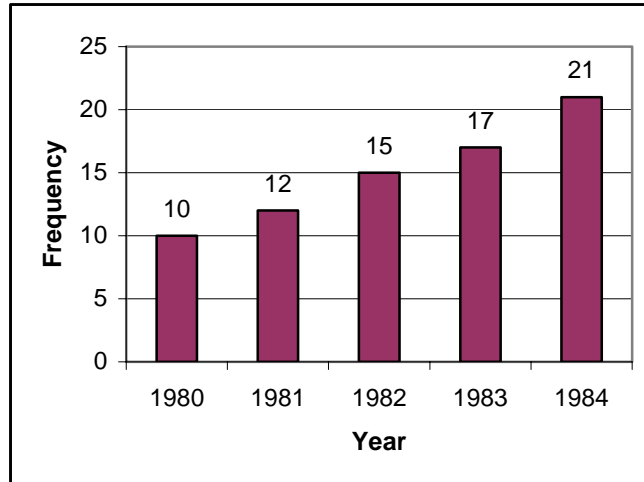


Fig.: Showing malignant obstructive jaundice prevalence rate/2000 population.

ii. **Subdivided bar diagram:** These diagrams are used to represent various parts of the total.

Example- Represent the following data by subdivided bar diagram.

The number of cancer patient admitted in three medical colleges hospital as follows.

Table-1

Year	RMCH	DMCH	BMCH
1983-84	70	86	50
1984-85	92	110	70
1985-86	56	90	71
1986-87	35	75	56
1987-88	78	87	52
1988-89	55	59	71
1989-90	99	112	72
1990-91	82	75	59

Solution: Since we have to show three different variable, subdivided bar diagram will be more appropriate.

Interpretive Phase

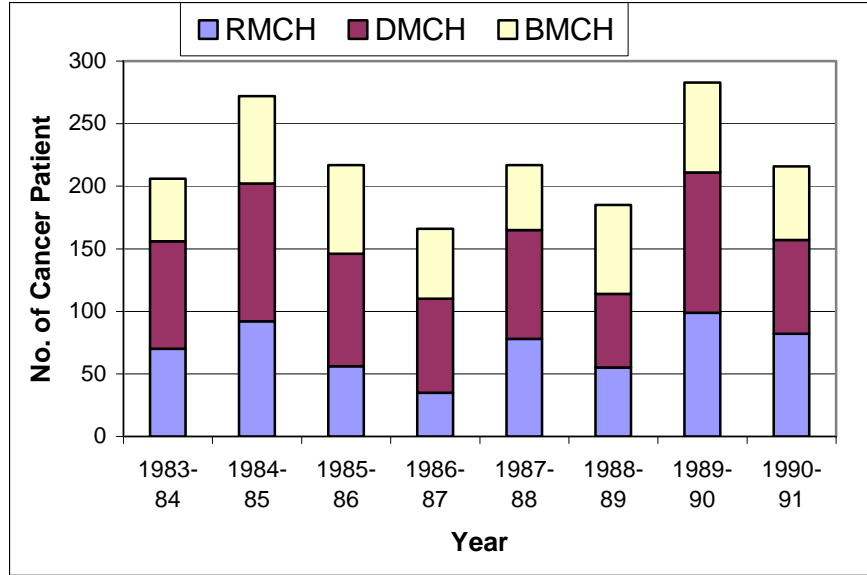


Fig.: Subdivided bar diagram.

iii. **Multiple bar diagram:** Two or three bars are drawn side by side to represent different phenomenon relating to the same period of time. A multiple bar diagram can be drawn from the data of table-1.

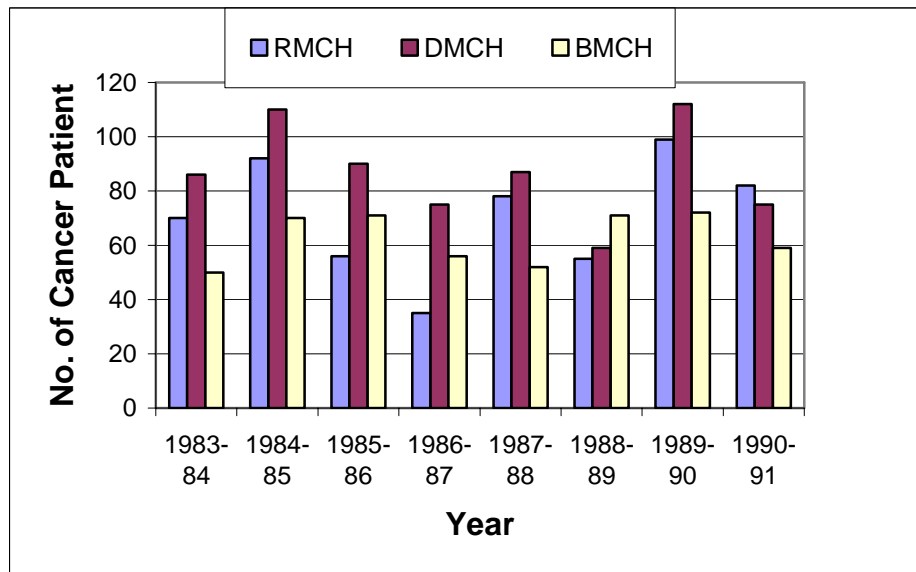


Fig.: Multiple bar diagram.

2. **Histogram:** It is a graph, which represents the frequencies corresponding to each class in a frequency distribution by vertical rectangles. The x-axis represents the class intervals with the breadth of

each column showing the magnitude of the class interval. The frequency of the class interval is shown by the height of the column along the y-axis.

Let us consider the following frequency distribution table-2.

Class interval	Mid value	Frequency	Cumulative frequencies
25-30	27.5	5	5
30-35	32.5	7	12
35-40	37.5	10	22
40-45	42.5	9	31
45-50	47.5	6	37

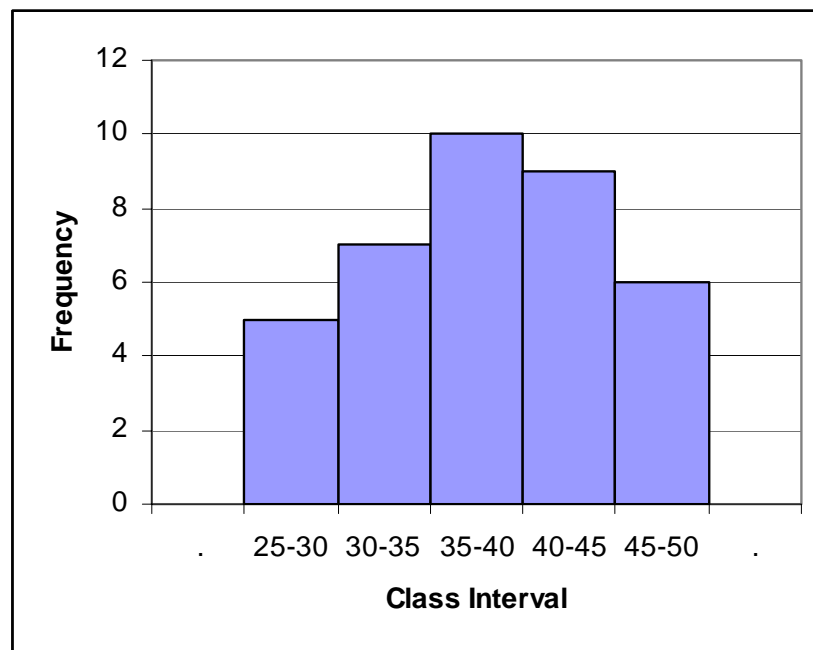


Fig.: Histogram

3. **Frequency polygon:** In frequency polygon the mid-values of the continuous class intervals are represented along x-axis and the frequencies corresponding to the class intervals are presented along the y-axis. The class frequencies are plotted against the mid-values of the respective class intervals. These points are then joined and the last points are then brought down at each end to the x-axis by joining it to the mid-values of the next out lying interval of zero frequency. The polygon thus obtained is called frequency polygon.

Interpretive Phase

Frequency polygon of table-2 is-

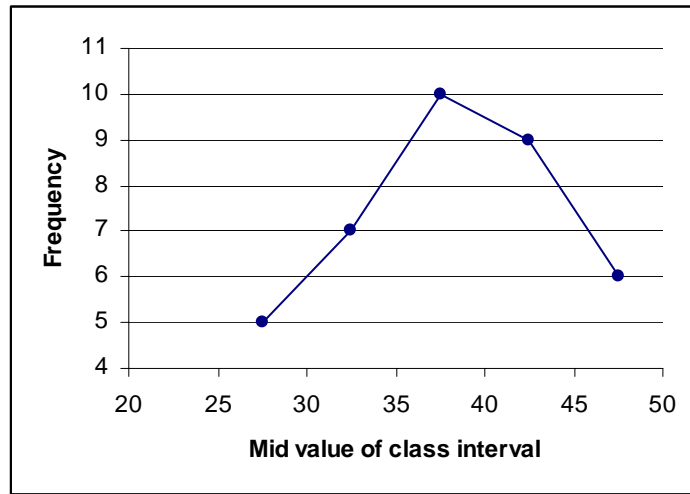


Fig.: Frequency polygon.

4. **Cumulative frequency polygon:** In cumulative frequency polygon the upper limits of the continuous class interval are represented in x-axis and cumulative frequencies are represented to y-axis. It is non-decreasing but may be parallel to x-axis.

From table-2 cumulative frequency polygon can be drawn as follow.

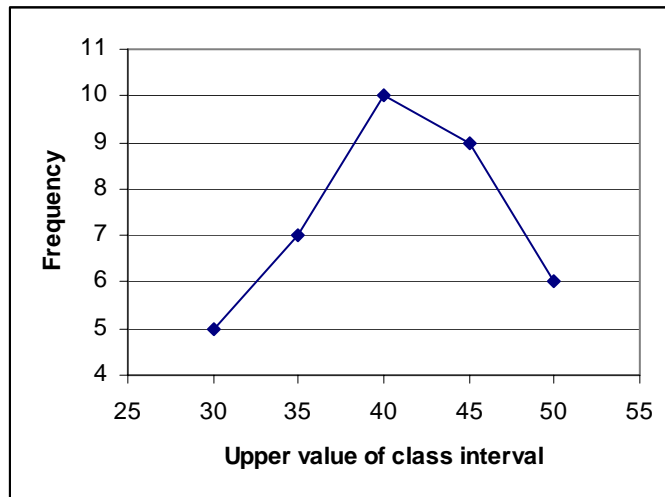


Fig.: Cumulative frequency polygon.

5. **Ogive curve:** A free hand curve to smooth a cumulative frequency polygon is called an ogive curve. From table-2 ogive curve is drawn as follow.

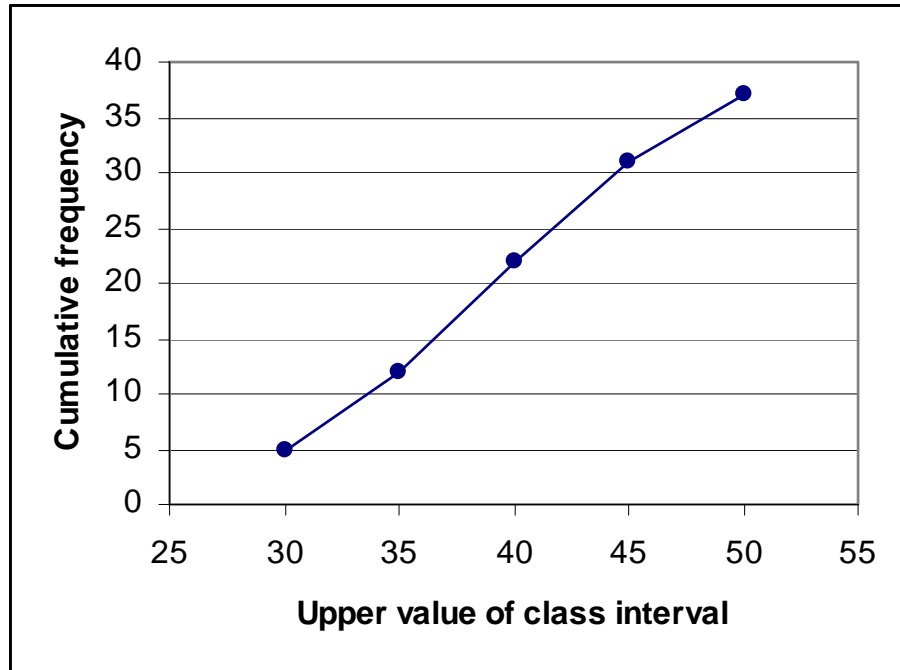


Fig.: Ogives curve.

6. **Pie diagram:** It is a circle divided into parts to reveal the various components of the data. Percentage of the frequency of each class interval is calculated at first by the following formula.

$$\frac{\text{Class frequency}}{\text{Total no.}} \times 100$$

From table-2, pie-diagram can be drawn as follows.

Class interval	% Frequency
25-30	13.51
30-35	18.91
35-40	27.02
40-45	24.32
40-45	16.21

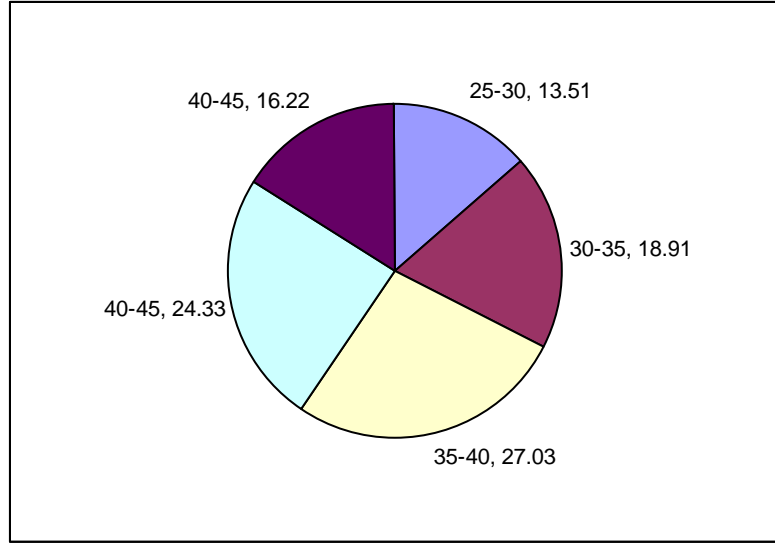


Fig. Pie-diagram.

1.6. Mean

Mean of set of observations is the sum of all observations divided by the number of observations e.g., the mean \bar{x} of n ungrouped observations $x_1, x_2, x_3, \dots, x_n$ is given by-

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

$$\therefore \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\text{or } \bar{x} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_n x_n}{f_1 + f_2 + \dots + f_n} \text{ (For group data)}$$

$$= \frac{\sum_{i=1}^n f_i x_i}{N}$$

$$\text{Where } \sum_{i=1}^n f_i = N$$

Example: \bar{x} of series 3, 5, 4, 7, 4, 2, 6, 3, 5, 3, 2, 4, 6, 5, 7 is.

$$\begin{aligned} \bar{x} &= \frac{3+5+4+7+4+2+6+3+5+3+2+4+6+5+7}{15} \\ &= \frac{66}{15} = 4.4 \qquad \therefore \bar{x} = 4.4 \end{aligned}$$

Example 2: Calculate mean from the frequency distribution of wages with class interval of two taka each from the following data of daily wages received by 35 labours in a certain factory.

Class interval of wages (Taka)	Numbers of labours (f_i)
11-13	3
13-15	4
15-17	5
17-19	10
19-21	6
21-23	4
23-25	3
Total	35

Solution

Class interval of wages (Taka)	Number of Labours (f_i)	Mid value of class interval (x_i)	$f_i x_i$
11-13	3	12	36
13-15	4	14	56
15-17	5	16	80
17-19	10	18	180
19-21	6	20	120
21-23	4	22	88
23-25	3	24	72
Total	35		632

$$\text{Mean, } \bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{632}{35} = 18.06 \text{ Tk(app)}$$

Merits

- i. It is rigidly defined and easy to calculate.
- ii. It is easy to understand and easy for algebraic treatment.
- iii. It takes all the observations into account.
- iv. It is less affected by sampling fluctuation.

Demerits

- i. It is affected by extreme values.

Interpretive Phase

- ii. It is impossible to calculate if the extreme classes of the frequency distribution are open.
- iii. The value of the mean may not occur in the series.
- iv. Cannot be located graphically.

1.7. Median

It is defined as the value of central or middle item, which divides series into two equal parts, when the series is arranged in descending or ascending order of magnitude.

For ungrouped data, when n is odd, the middle observation i.e., the $(\frac{n+1}{2})$ th observation will be the median in the series.

Again when n is even, the median will be the mean of $\frac{n}{2}$ th and $(\frac{n}{2}+1)$ th observations in the series.

For examples, the median of the observations 5, 10, 7, 3, 2, i.e., 10, 7, 5, 3, 2, is 5 and the median of 11, 3, 9, 5, 7, 13, i.e., 13, 11, 9, 7, 5, 3, is $\frac{9+7}{2} = 8$.

For grouped frequency distribution the median is given by-

$$Me = L + \frac{\frac{N}{2} - F}{f} \times C$$

Where-

L = The lower limit of the median class (median class is that class which contains $\frac{n}{2}$ th observation of the series).

N = Total number observation.

F = Cumulative frequency of the class just preceding the median class.

f = Frequency of the median class and

C = Length of the median class.

Example- Find out median from the following frequency distribution table-3.

Class interval of wages (Taka)	Number of labours (f _i)	Cumulative frequency
11-13	3	3
13-15	4	7
15-17	5	12
17-19	10	22
19-21	6	28
21-23	4	32
23-25	3	35

$$\therefore \text{Me} = L + \frac{\frac{N}{2} - F}{f} \times C$$

$$= 17 + \frac{\frac{35}{2} - 12}{10} \times 2 \text{ (Putting all values)}$$

Here,
L = 17
N = 35
F = 12
f = 10
C = 2

Merits

- i. Median is rigidly defined.
- ii. It is easily understood and easy to calculate.
- iii. It is not all affected by extreme values.
- iv. It can be calculated from frequency distribution with open end.

Demerits

- i. In case of even number of observations, median cannot be defined exactly.
- ii. It is not based on all the observations.
- iii. It is not easy for algebraic treatment.
- iv. It is affected much by sampling fluctuation

1.8. Mode

The mode is that observation of the variable for which the frequency is maximum.

For example: The mode of the observations 2, 5, 9, 5, 8, 7, 4, is 5.

For grouped frequency distribution the mode is given by

Interpretive Phase

$$M_o = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C$$

Where-

L = The lower limit of the modal class (modal class is that class for which the frequency is maximum)

Δ_1 = The difference between the frequency of the modal class and pre-modes class.

Δ_2 = The differences between the frequency of the modal class and post-modal class and C = Length of the modal class.

Example- Find out mode from the frequency distribution given in table-3.

Solution: Here the modal class is (17-19) because in that class the frequency is maximum i.e., 10.

$$\begin{aligned} M_o &= L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C \\ &= 17 + \frac{5}{5+4} \times 2 = 18.11(\text{Tk}) \end{aligned}$$

Here-

$$\begin{aligned} L &= 17 \\ \Delta_1 &= 5 \\ \Delta_2 &= 4 \\ C &= 2 \end{aligned}$$

Merits

- i. Mode is easy to understand and easy to calculate.
- ii. It is not at all affected by extreme values
- iii. It can be calculated from frequency distribution with open class.

Demerits

- i. Mode is not clearly defined in case of bi-modal or multimodal distribution.
- ii. It is not based on all observations.
- iii. It is difficult for algebraic treatments.
- iv. It is affected to a great extent by sampling fluctuation.

1.9. Standard Deviation

Standard deviation of a set of observation is the square root of the arithmetic mean of the squares of the deviations of the arithmetic mean from the given observations.

Let $x_1, x_2, x_3 \dots x_n$ be any set of n observations and their corresponding frequencies $f_1, f_2, \dots f_n$ respectively and their arithmetic mean is \bar{x} , then the standard deviations is-

$$S.D = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \text{ (For ungrouped data)}$$

$$\text{and } S.D = \sqrt{\frac{\sum_{i=1}^n f_i (x_i - \bar{x})^2}{N}} \text{ (For grouped data)}$$

Example (For ungrouped data)- Find the standard deviation from average two days wages of ten workers working in a hospital.

Sl. No	Wages (Tk)	Sl. No	Wages (Tk)
1	320	6	340
2	310	7	325
3	315	8	321
4	322	9	320
5	326	10	331

Solution: Calculation of standard deviation.

Sl. No	Wages	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	320	-3	9
2	310	-13	169
3	315	-8	64
4	322	-1	1
5	326	+3	9
6	340	+17	289
7	325	+2	4
8	321	-2	4
9	320	-3	9
10	331	+8	64
n=10	$\sum x_i=3230$	$\sum (x_i - \bar{x})=0$	$\sum (x_i - \bar{x})^2=622$

$$S.D = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

$$= \sqrt{\frac{622}{10}}$$

$$\therefore S.D = 7.89(\text{Ans})$$

Where

$$n=10$$

$$\sum (x_i - \bar{x})^2 = 622$$

Interpretive Phase

Merits

- i. It is rigidly defined.
- ii. It takes all the observations into account.
- iii. It is amenable to algebraic treatments.
- iv. It is less affected by sampling fluctuation.
- v. The standard deviation of the combined services can be obtained if the means, standard deviations and number of observations in each series are given.

Demerits

- i. It is affected by extreme values.
- ii. It cannot be calculated if the extreme classes open.
- iii. The frequency distributions are open.



1.10. Exercise

1.10.1. Tick (✓) the correct answer

1. Which is not mean
 - a. Arithmetic mean
 - b. Median
 - c. Geometric mean
 - d. Harmonic mean
2. Statistics means
 - a. Political state
 - b. Social state
 - c. Moral state
 - d. Total calculation

1.10.2. Determine True or False

- a. Statistic means social state
- b. Variable always remain fix
- c. Percentile divide many parts of a series of observations
- d. Simple bar diagram represents one variable.

1.10.3. Fill in the blanks

- a. Number of events in each group is called
- b. Subdivided bar diagram represents of the total.
- c. The mode is that observation of the variable for which the frequency is
- d. Frequency distribution describes the patterns of through its range.

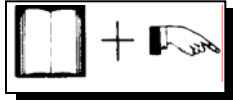
1.10.4. Match (Match column A with column B)

Column A	Column B
1. Anything that varies with the limits is called	i. variable
2. Variance is the estimation of the variation	ii. of the sample
3. Sum of all observation divided by the number of observation is	iii. mean
4. Median is the value of central or middle	iv. item
5. Histogram represents the frequency to	v. each class

1.10.5. Short questions

- 1. Define statistics.
- 2. Discuss the common tools of statistics.
- 3. Define mean, median, mode and standard deviation.
- 4. Calculate, mean, median and mode from the following data.
17, 34, 32, 54, 35, 84, 86, 14, 35, 84, 35, and 87.

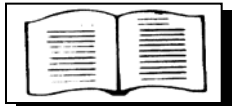
Lesson 2: Processing, Analysis and Interpretation of Data



2.1. Learning Objectives

On completion of this lesson you will be able to-

- know about data processing
- learn how to analyze data
- interpret data.



2.2. Introduction

The data, after collection, has to be processed, analyzed in accordance with the outline laid down for the purpose at the time of developing the research plan. It may be fair to say that research consists in general of two larger steps- the gathering data, and analysis of these data, but no amount of analysis can validly extract from the data factors which are not present. To avoid making conclusions or interpretations from insufficient or invalid data, the final analysis must be anticipated in detail, when plans are being made for collecting information. The problem should be analyzed in detail to see that what data are necessary in its solution and to be assured that the method used will provide definite answers.

2.3. Steps in Data Processing

After carefully collecting data from the study subjects, the investigator can begin a series of important processes to prepare the data for analysis. Technically speaking processing implies editing, coding, classification and tabulation of collected data so that they are amenable to analysis.

1. Editing: Editing of data is a process of examining the collected raw data (specially in surveys) to detect errors and omissions and to correct these when possible. As a matter of fact, editing involves of careful scrutiny of the completed questionnaires and/or schedules.

Editing is done to assure that the data are accurate, consistent with other facts gathered, uniformly entered, as complete as possible. Editing may be of two types-

- Field Editing:** Field editing is done in the review of the reporting forms by the investigator at the time of recording the respondent's responses. This sort of editing should be done as soon as possible after the interview, preferably on the very day or on the next day.

- ii. **Central Editing:** Central editing is taken place when all forms or schedules have been completed and returned to the office. In this case, a single editor takes a thorough editing in a small study and a team of editors in case of a large inquiry.

2. Coding: Coding refers to the process of assigning numerals or other symbols to answer so that responses can be put into a limited number of categories. Data should be coded in such, as a way that a specific answer can be placed in one cell in a given category set. Coding decisions should usually be taken at the designing stage of the questionnaire. Code is an abbreviation a number or an alphabet, which is assigned by the researcher to every schedule item and response category. Coding for computer analysis frequently involves transforming phrases or words to numbers.

3. Classification: Classifying refers to the dividing up of the data into different categories, classification, or sub-headings for use. Most research studies result in a large volume of raw data, which must be reduced into homogeneous groups if we are to get meaningful relationships. Data having a common characteristic are placed in one class and in this way the entire data get divided into a number of groups or classes. Classification can be one of the following two types, depending upon the nature of the phenomenon involved.

a. Classification according to attributes: Descriptive data (such as literacy, sex, honesty, etc.) refer to quality not quantity. When the data are obtained on the basis of certain attributes is known as statistics of attributes and their classification is said to be classification according to attributes.

Such classification can be simple classification or manifold classification. In simple classification we consider only one attribute and divide the universe into two classes-one classes consisting of items possessing the given attribute and the other class consisting of items, which do not possess the given attribute. But in manifold classification we consider two or more attributes simultaneously.

b. Classification according to class-intervals: Numerical data (such as weight, height, income etc.) refer to quantity not quality. This data can be measured through some statistical units. Data relating to income, production, age, weight, etc. come under this category. Such data are known as statistics of variables and classified on the basis of class intervals. For instance, persons whose incomes, say, are within Tk. 201 to Tk. 400 can form one group, those whose incomes are within Tk. 401 to Tk. 600 can form another group and so on. In this way the entire data may be divided into a number of groups or classes which are usually called “class-interval”. Each group or class-interval, thus, has an upper limit as

Interpretive Phase

well as lower limit, which are known upper and lower limit respectively. The number of items, which fall in a given class, is known as the frequency of the given class.

4. Tabulation: When a mass of data has been assembled, it becomes necessary for the researcher to arrange the same in some kind of concise and logical order. This procedure is referred to as tabulation. Thus, tabulation is the process of summarizing raw data and displaying the same in compact form (i.e., in the form of statistical table) for further analysis. In a broad sense, tabulation is an orderly arrangement of data in columns and rows.

Tabulation is essential because of the following reasons.

- i. It conserves space and reduces explanatory and descriptive statement to a minimum.
- ii. It facilitates the process of comparison.
- iii. It facilitates the summation of items and the detection of errors and omission.
- iv. It provides a basis for various statistical computations.

Tabulation can be done by hand or by mechanical or electronic devices. The choice depends on the size and type of study, cost considerations, time pressures and the availability of tabulating machines or computers.

Tabulation may be classified as simple and complex tabulation. The former type of tabulation gives information about one or more groups of independent questions, where as the later type of tabulation shows the division of data in two or more categories and as such is designed to give in formation concerning one or more sets of inter-related question.

2.4. Analysis of Data

Term analysis refers to the computation of certain measures along with searching for patterns of relationship that exist among data-groups. Analysis, particularly in case of survey or experimental data, involves estimating the values of unknown parameters of the population and testing of hypotheses for drawing inferences. Analysis may, therefore, be categorized as-

- i. Descriptive analysis
- ii. Inferential analysis (often known as statistical analysis).

Types of analysis-

- i. Descriptive analysis- Descriptive analysis is largely the study of distributions of one variable. It may be of following types-
 - a. Unidimensional analysis- It describes one variable.
 - b. Bivariate analysis- It describes two variables.
 - c. Multivariate analysis- It describes more than two variables. The following analyses are involved in multivariate analysis.
 1. Multiple regression analysis- This analysis is adopted when the research has one dependent variable which is presumed to be a function of two or more independent variables. The objective of this analysis is to make a prediction about variable based on its covariance with all the concerned independent variables.
 2. Multiple discriminant analysis- This analysis is appropriate when the research has a single dependent variable that cannot be measured, but can be classified into two or more groups on the basis of some attribute.
 3. Multivariate analysis of variance (or multi-ANOVA)- This analysis is an extension of two-way ANOVA, where in the ratio of among group variance to within group variance is worked out on a set of variables.
 4. Canonical analysis- This analysis can be used in case of both measurable and non-measurable for the purpose of simultaneously predicting a set of dependent variables from their joint covariance with a set of independent variables.
- ii. Correlation analysis- Correlation analysis studies the joint variation of two or more variables. For determining the amount of correlation between two or more variables.
- iii. Causal analysis- Causal analysis is concerned with the study of how one or more variables affect changes in another variable.

Planning data analysis required the investigator to consider the domain of descriptive and inferential statistics. The statistical procedures are used to give organization and meaning of data. Procedure that allows researchers to describe and summarize data is known as descriptive statistics. Procedures that allow researchers to estimate how reliably they can make predictions and generalize finding based on the data are known as 'inferential statistics'. In other descriptive, statistics summarize information about a sample, whereas inferential statistics allow the researcher to make conclusions that extend beyond the sample studied to the population of interest.

2.5. Inferential Statistics

Inferential statistics combine mathematical processes and logic that allows researchers to test hypotheses about a population using data obtained from probability samples. Inferential statistics enables to go beyond the immediate description of the results of individual research studies in ways that provide the best possible bases for clinical practice or further research. The most commonly used purpose of inferential statistics is hypothesis testing. Statistical hypothesis testing allows researcher to make objective decision about the out come of their study. The null hypothesis, which is the hypothesis that actually can be tested by statistical methods, would state that there is no difference between groups.

Basically these statistical tests are two types of hypothesis-

- a. That there is difference between groups i.e., test of difference.
- b. That there is a relationship between two or more variables i.e., test of relationship or association.

Some null hypothesis tests with example are given below.

Null hypothesis- The hypothesis of no difference. No difference in result between the experimented and control groups.

If there is a difference the hypothesis is rejected. Null hypothesis is proved. Then both groups belong to same class.

Test for hypothesis

- i. Student's test.
- ii. Chi square test (χ^2) test
- iii. 'z' test
- iv. Correlation coefficient test ('r' test)
- v. Analysis of variance ('F' test (ANOVA))

1. Students 't' test: It is done to test the difference between two groups not within the group. Note- The minimum No. of subject in any group in an experiment is 6 (depending upon the type of experiment).

$$t = \frac{\text{Difference of means } (m_1 - m_2)}{\sqrt{(SE_1)^2 + (SE_2)^2}} \quad \begin{array}{l} \text{Here, } m_1 = \text{Mean of group - I} \\ m_2 = \text{Mean of group - II} \end{array}$$

Degrees of freedom: It indicates number of independent observation.

$Df=(n_1-1)+(n_2-1)$ Here n_1 = no of observation in group- I
 n_2 = no of observation in group- II

Example:

Group-I	Group- II
80	125
85	130
75	140
84	145
83	150
$n_1=5, m_1=81.4$	$n_2=5, m_2=138$

In group-I SE_1

x	x̄	(x-x̄)	(x-x̄) ²	Σ(x-x̄) ²	SE ₁
80	81.4	-1.4	1.96	65.2	1.80
85	81.4	3.6	12.96		
75	81.4	-6.4	40.96		
84	81.4	2.6	6.76		
83	81.4	1.6	2.56		

$$SE_1 = \sqrt{\frac{\sum(x-\bar{x})^2}{n(n-1)}} = \sqrt{\frac{65.2}{5(5-1)}} = \sqrt{\frac{65.2}{20}} = \sqrt{3.26} = 1.80$$

In Group II. SE_2

x	x̄	(x-x̄)	(x-x̄) ²	Σ(x-x̄) ²	SE ₂
125	138	-13	169	430	4.63
130	138	-8	64		
140	138	2	4		
145	138	7	49		
150	138	12	144		

$$SE_1 = \sqrt{\frac{430}{5(5-1)}} = \sqrt{\frac{430}{20}} = 4.63$$

$$t = \frac{81.4 - 138}{\sqrt{(1.80)^2 + (4.63)^2}} = \frac{56.6}{\sqrt{3.24 + 21.43}} = \frac{56.6}{\sqrt{24.67}} = 56.6 / 4.96 = 11.41$$

t = 11.41. Now if we take the level of significance as 0.01

Interpretive Phase

The table value of “t” in level. 0.01 against df = 8 is

(5.04) the result is highly significant and null hypothesis is rejected.

1. CHI-SQUARE (χ^2) TEST: The calculation of quantity called chi-square, from Greek word “chi” (χ^2) and pronounced as “Key”. The chi-squatre test is done to show association between observed and expected value.

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

O = Observed value
E = Expected value

Especially utilized in assessing the effectiveness of a new vaccine of drugs.

May be used to compare effectiveness between two drugs.

Problem

Three drugs A, B and C were given in a clinical trial among three hundred (300) patients, of these patients 40 were exposed to drug A, 100 to drug B and 160 to drug C respectively. The cure rates of these drugs were 20, 60 and 120 for drugs A, B and C respectively. Determine the effectiveness of these drugs.

$$E = \frac{\text{Row total X column total}}{\text{Grand total}}$$

Treatment group	Cured	Not cured	Row total
A	20 (26.67)	20 (13.33)	40
B	60 (66.67)	40 (33.33)	100
C	120 (106.67)	40 (53.33)	160
Column total	200	100	300

(Grand total)

Expected values of the different groups are as follows.

$$\text{Group A: Cured} = \frac{40 \times 200}{300} = 26.67$$

$$\text{Not cured: } 40 \times 100 / 300 = 13.33$$

$$\text{Gr. B: Cured} = 100 \times 200 / 300 = 66.67$$

$$\text{Not cured} = 100 \times 100 / 300 = 33.33$$

$$\text{Gr. C: Cured} = 160 \times 200 / 300 = 106.67$$

$$\text{Not cured} = 160 \times 100 / 300 = 53.33$$

O	E	O-E	$\sum \frac{(O-E)^2}{E}$	df
20	26.67	-6.67	1.67	2
20	13.33	+6.67	3.34	
60	66.67	-6.67	0.67	
40	33.33	+6.67	1.33	
120	106.67	+13.33	1.67	
40	53.33	-13.33	3.33	
$\chi^2 =$			12.01	

$$\text{df} = (\text{Row}-1) \times (\text{column}-1)$$

$$= (3-1) \times (2-1) = 2 \times 1 = 2$$

$$\chi^2 = 12.01. \text{ df} = 2. p < 0.01$$

The result is significant and null hypothesis is rejected.

2.6. Interpretation of Data

By interpretation of data we mean the task of drawing conclusions or, inferences and of explaining their significance, after careful analysis of the collected data. The interpretation of research data cannot be considered in the abstract. In view of the diversity of research methods use in education, and the corresponding diversity of data, they seek; the interpretation of such data is best considered within context of each method. The analysis and interpretation of historical data. For example, best viewed in the light of historical method, its objectives and its limitations.

The process of interpretation is essentially one of stating that what the findings show. The findings of the study are the results, conclusions, interpretations, recommendations, generalizations, implications, future research and nursing practices. Interpreting the findings of a study involves a search for their meaning in relation to the problem, conceptual framework, purpose and all the research decisions made in developing and implementing the empirical phase of the study.

The interpretation of finding is an extremely important aspect of conducting a study. Interpreting the results of a study requires confrontation with three types of validity, i.e., explanatory validity,

Interpretive Phase

ecological validity and methodological validity. Explanatory Validity refers to that which the concept is chosen to account for that study findings to do so. This required examination of alternative, equally plausible explanation for the findings. Ecological validity refers to the extent to which the sample of observations in the study represents the substantive domain, the adequacy of the relationships between the study design and substance being studied. Methodological validity refers to the degree of which the findings are a function of set or methods used to test the theory (Grinberg and McGrath 1982).

Interpretation is by no means a mechanical process. It calls for critical examination of the results of one's analysis in the light of all the limitations of his data gathering. It is a very important step in the total procedure of research. It is important to recognize that errors can be made in the interpretation. Just as that can in any of other steps of the scientific method and the specific errors to be guarded against vary with the different research methods. This steps is purely subjective and many errors are made at this point. It however, one is careful and critical of his own thinking, he should be able to make satisfactory interpretations.

The following are among the more common errors of interpretation, which need to be avoided.

Failure to see the Problem in Proper Perspective

Sometimes, investigator may have an inadequate grasp of the problem in its broad sense and too close a focus in its immediate aspect.

Failure to Appreciate the Relevance of Various Elements

The investigator may fail to see the relevance of the various elements of the situation due to an inadequate grasp of the problem, too rigid a mind set or even a Jack of imagination. This may cause the investigator to overlook the operation of a significant factor. Consequently, the outcomes of the study are attributed to the wrong antecedent.

Failure to Recognize Limitations in the Research Evidence

These limitations may be of many types of such a non-representativeness in sampling, biases in the data, inadequacies in the research design, defective, data-gathering instrument and inaccurate analysis.

Faulty inferences on the basis of inadequate data. Errors in the use of various tools of analysis like mean, median, mode, percentage.

Faulty generalizations on the basis of incorrect and unrepresentative sample.

Ignoring Selective Factors

In investigations where a selective group is made the subject of a study or where a selective factor is operating on the situation studied one is likely to reach unwarranted conclusions if one ignores the selective factors.

Difficulties of interpretative evaluation
Inappropriate comparisons.



2.7. Exercise

2.7.1. Tick (✓) the correct answer

1. Data processing includes
 - a. Editing
 - b. Tabulation
 - c. Coding
 - d. All above.

2. In central tendency the data are cluster
 - a. At right side
 - b. At middle
 - c. Left side
 - d. None of the above.

2.7.2. Determine True of False

1. Data editing is necessary for detecting error
2. Tabulation cannot be done by electronic devices
3. Descriptive analysis is done for one variable
4. A frequency distribution enlist a series of observations of a variable

2.7.3. Fill in the blanks

1. The process of assigning numerals is called
2. Bivariate analysis describes
3. Range is equal to the highest value minus
4. The relation between two variables is called

Interpretive Phase

2.7.4. Matching (Match column A with column B)

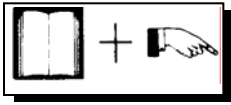
Column A	Column B
1. Descriptive data refer quality not	i. quality
2. Unidimensional analysis describes	ii. one variable
3. Range is mostly used as a rough measure	iii. of variability
4. Degree of freedom indicates numbers of	iv. independent observation

2.7.5. Short questions

1. Define data processing.
2. Discuss about descriptive statistics.
3. Write down about inferential statistics.

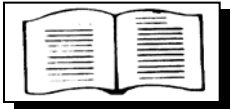
Lesson 3: Writing the Research Report

3.1. Learning Objectives



On completion of this lesson you will be able to-

- to learn how to write the research report
- to know guidelines for writing research report
- to know about the different format of the research report



3.2. Writing the Research Report

Writing the research report is the highlight of the research project. The research task is not completed until the report has been written. A research report is the ultimate output of the research process. A good research report should possess following characteristics.

- Conciseness- Lack of redundancy, conciseness saves the readers time and forces the writer to refine his ideas.
- Clarity- Lucidity of expression, clarity helps the reader to understand the points being made.
- Honesty- Freedom from fraud. Honesty is necessary to maintain the respect of the reader and the integrity of the author.
- Completeness- Full details, completeness results in a full report, thus enabling the reader to evaluate the study.
- Accuracy- Quality of being precise and free of error. Accuracy is necessary to maintain the integrity of the author in the eyes of the reader.

3.3. Guidelines for Writing Research Report

There are certain important guidelines to be followed while writing a research report as follows-

- i. Develop your thinking- There is a high positive correlation between good thinking and effective writing. Research is not merely the accumulation, evaluation and assimilation; it is a process of rebuilding facts into a meaningful whole.
- ii. Divide your narration- Divide your narration into paragraphs and use of informative headlines whenever necessary. Paragraphing is an important feature of any report. It groups sentences around one central thought or idea. Linking one paragraph to another is an essential technique for maintaining continuity.

Interpretive Phase

- iii. Use present tense and active voice- Results of research should be valid in the present. Otherwise, there is no use to present them.
- iv. Minimize the use of technical language or jargon- Most disciples are criticized for their use of technical jargon. They stress here, the range of possible readers can be increased by use of simple straightforward language. Clarity, conciseness, and simplicity are critical attributes of any kind of good writing.
- v. Use visual aids- Use visual aids in the form of tables and figures to illustrate the principal findings of the study. It is important that such illustrations be used to emphasize points made in the next rather than to replace them.
- vi. Be objective- The report should be unbiased and objective justified by facts. All references to the thoughts and works of others should, however, be properly acknowledged in footnotes. Otherwise, the writer would be considered guilty of plagiarism.
- vii. Treat data confidentially- Confidentiality is not an issue if individuals have voluntarily provided data with full awareness that these will be revealed to others. But where the researcher has promised to the respondents to protect that anonymity, the same should be written in such a way as to preclude the possibility of respondents' identification.
- viii. Revise and rewrite- Revising is part of writing. Few writers are so expert that they can produce what they are on the first try.

3.4. Different Steps in Writing Report

Research reports are the product of slow, painstaking, accurate inductive work. The usual steps involved in writing report are-

- a. Logical analysis of the subject matters.
- b. Preparation of the final outline.
- c. Preparation of rough draft.
- d. Rewriting and polishing.
- e. Preparation of the final bibliography.
- f. Writing the final draft.

A brief description of the above steps are given below-

- a. **Logical analysis of the subject matter:** It is the first step, which is primarily concerned with the development of a subject. There are two ways in which to develop a subject- (i) Logically and (ii) chronologically. The logical development is made on the basis of mental connections and association between the one thing and another

by means of analysis chronological development is based on a connection or sequence in time or occurrence. The directions for doing or making something usually follow the chronological order.

- b. **Preparation of the final outline:** It is the next step in writing the research report. Outlines are the framework upon which long written works are constructed. They are an aid to the logical organization of material and reminder of the points to be stressed in the report.
- c. **Preparation of the rough draft:** This follows the logical analysis of the subject and preparation of the subject and preparation of the final outline. Such a step is important for the researcher now sits to write down what he does in the context of his research study.
- d. **Rewriting and polishing of the rough draft:** This step is the most difficult part of all formal writing. Usually this step requires more time than the writing of the rough draft. The careful revision makes difference between a mediocre and a good piece of writing. While rewriting and polishing, one should check the report for weaknesses in logical development or presentation.
- e. **Preparation of the final bibliography:** Next in order comes the task of the preparation of the final bibliography. The bibliography, which is generally appended to the research report, is a list of books in some way pertinent to the research, which has been done.
- f. **Writing the final draft:** The final draft should be written in a concise and objective style and in simple language, avoiding vague expression such as “it seems” “there may be”, and the like ones. A research report should not be dull, but must enthuse people and maintain interest and must show originality.

3.5. Format of the Research Report

The report of a research not only serves to record and communicate the procedure and the results but it also fulfils an important function in the process of research. A research report usually follows fairly standardized pattern. The following outlines research report or formal were establish in some institutions as follows.

A research report usually follows fairly standardized pattern. The following outlines research report of format were established in some institutions as follows-

Format of Research Report

1. a. Introduction
- b. Importance of the general problem

Interpretive Phase

- c. Specific problem
2. Research statement
 - a. Statement of what the researcher studied and how
 - b. Definitions of concepts/ variable
 - c. Hypothesis, if any
 - d. Objectives of the study: purposes
 - e. Ethical implications of the research.
3. Review of the literature
 - a. Review of related and competing theories
 - b. Review of relevant research
 - c. Specification of theory and research used in the study
 - d. Review of observations to formulate theory if applicable
4. Research design
 - a. Description of the particular design used
 - b. Description of the control used, if any
 - c. Discussion of the validity and reliability of the design
5. Sampling
 - a. Description of the target population
 - i. Discussion of how the sample was drawn, sample size response rate
 - ii. Discussion of bias, if any
6. Methodology
 - a. Description of the method of data collection used
 - b. Description of categories, instruments, scales, operational definitions
 - c. Discussion of the reliability and validity of methods, instruments
7. Pilot study
 - a. Description of finding
 - b. Subsequent revisions

8. Analysis and interpretation of data
 - a. Description of statistics used and how the data was analyzed
 - b. Summary of data in graphs and tables, with narrative explanation
 - c. Interpretation of findings
9. Conclusion
 - a. Implication of findings for nursing
 - b. Recommendations, suggestions for future research
10. Bibliography and appendices
11. Abstract

Contents of the Format

Whatever the format, the research uses, he/she should maintain proper sequence in his/her presentation. The following outline presents the usual sequence of various sections.

Preliminary sections or front matter

1. Title page
2. Approval sheet
3. Acknowledgement (if any)
4. Preface or foreword
5. Table of contents
6. List of tables (if any)
7. List of figures (if any)

Main body of the report

1. Introduction
 - a. Statement of the problem
 - b. Significance of the problem
 - c. Purpose of the study
 - d. Assumptions and delimitations
 - e. Definition of important terms
 - f. Statement of hypothesis

Interpretive Phase

- g. Assumption underlying hypothesis
2. Review of related literature or analysis of previous research
3. Design of the study
 - a. Procedures used
 - b. Methods of gathering data
 - c. Description of data gathering instrument
4. Presentation and analysis of data
 - a. Text
 - b. Tables
 - c. Figures
5. Summary and conclusions
 - a. Brief restatement of the problem and procedures
 - b. Description of procedures used
 - c. Principal findings and conclusions
 - d. Recommendations for further research.

A. Preliminary Section

1. Title page- The first page of the report is the title page. It should indicate the purpose of the research, its completion date, for whom it was conducted any by whom. If the research is of a confidential in nature, this should be stated along with a list of individuals who should be allowed to see it. Usually the title page includes the following.
 - a. The name of the topic or the title of the study
 - b. Full name of the candidate and his previous academic background.
 - c. Name the faculty/department and institution to which the report is submitted.
 - d. Mention the degree for which the report is presented.
 - e. The date of submission.
2. Approval sheet- If the institution requires an approved sheet, a page of the dissertation/thesis allots space for the necessary certificate and the signature of the supervisor/advisor or the chairman/the member of the committee or Head of the Department of concerned.
3. Acknowledgements- The acknowledgement page is largely a matter of courtesy the investigator acknowledges the guidance and assurance he

has received in carrying out study. Good taste calls for acknowledgements to be expressed in a simple and restrained language

4. Preface or foreword- Sometimes preface or foreword, one or two pages long follows the acknowledgement page, containing some initial remarks and perhaps a brief statement of the scope, aim and general character of the research.
5. Tables of contents- The table of contents indicates the organism of the report by listing the various chapters and their page numbers, so that the decision maker or any body interested in reading the report can easily locate the required information in the report.

All the pages in the preliminary section are numbered at the center of the bottom margin with lower case Roman numerals (i, ii, iii, iv-----).

B. Main Text

The main text provides the complete outline of the research report along with all details. Title of the research study is repeated at the top of the first page of the main text and then follows the other details on pages numbered consecutively, beginning with the second page. Each main text of the report should have the following sections.

1. Introduction: The purpose of introduction is to introduce the research project to the readers. It should contain a clear statement of the objectives of research i.e., enough background should be given to make clear to the reader why the problem was considered worth investigating.
2. Methods and Material: A description of methodology tells the reader about the technical aspects of the study. It explains the design of the study in detail, the size and selection of the sample, place and duration of work with dates, the variables and controls employed, the sources of data, the tools and methods of gathering data, the reliability of the instruments selected or constructed and the statistical procedures used in the analysis are carefully described. This section gives an accurate detailed description of how the work was done.
3. Presentation and analysis of data: The data analysis and interpretation may be presented in separate chapters or may be interpreted and presented in one or more chapters. This is the heart of the research report. Sometimes separate chapters are developed to the tabulation, analysis and the interpretation of data. The arrangement depends on the quantity of the study. Analysis and interpretation of data have to be done through the media of text tables and figures.

Interpretive Phase

4. Implications of the results: Toward the end of the main text, the research should again put down the results of his research clearly and precisely.

Such implications may have three aspects as stated below-

- a. A statement of the inferences drawn from the present, which may be expected to apply in similar circumstances.
 - b. The conditions of the present study, which may limit the extent of legitimate generalizations of the inferences drawn from the study.
 - c. The relevant questions that still remain unanswered or new questions raised by the study along with suggestions for the kind of research that would provide answers for them.
- vi. **Summary:** This section may be quite short in comparison to the body of the report. The summary includes brief restatement of the problem, description of procedures used, major findings and conclusions and recommendation for further research.

C. Reference Section

References at the end of the paper should be listed numerically alphabetically and in the same order that they have been cited in the text. Reference list should identify references cited (e.g., book, journal article, pamphlet, internet site, cassette tape or film) in sufficient detail so that others locate easily.

This section includes

1. Bibliography: A bibliography is list of sources used in the report. It contains the list of authors cited alphabetically by last names. Some bibliographies classify entries under such as books, periodical, newspapers reports, public documents.

Book

- i. Park J.E and K.Park. Textbook of Preventive and social medicine, XI Edn. Jabalpur (INDIA), M/s Banarsidas Bhonot, 1986.
- ii. Basavanthappa B.T. "Community Health Nursing" First Edn, Japee Brothers New Delhi, 1997

Journal Paper

- i. Hadjibabaie, M. N. Rastkari, A. Rezie and M. Abdollahi, 2005. The adverse drug reaction in the gastrointestinal tract: An overview Intl. J. Pharmacol 1(1):1-8.

Chapters in Book

- i. Bray R. A., 1994 the leucaena psyllid. In: Forage Tree legumes in Tropical Agriculture (edns R.C. Gutierrez and H.M., Shelton) pp.283-291. CAB International, Oxford.
2. **Appendix:** An appendix it included, follow the bibliography. It is the final section of the report. Items that will appeal to only a few readers or that may be needed only for occasional reference should be confined to an appendix.
3. **Index:** If a study is complex, of major importance or to be published in book or monograph form, it deals to prepare an alphabetized index, which follows the appendix.

3.5. Mechanics of Writing a Research Report

There are very definite and set rules which should be followed in the actual preparation of the research report or paper. The criteria of format should be decided as soon as the materials for the research paper have been assembled. The following points deserve mention so far as the mechanics of writing a report are concerned.

1. **Size and Physical Design:** The manuscript should be written on unruled paper $8\frac{1}{2} \times 11$ " in size. It is to be written by hand, then black or blue-back ink should be used. A margin of at least one and one-half inches should be allowed at the left hand and of at least half an inch at the right hand of the paper. There should also be one-inch margins, top and bottom. The entire manuscript, including reference, should be type double-spaced.
2. **Procedure:** Various steps in writing the report should be strictly adhered (as described earlier in this lesson).
3. **Treatment of quotations:** Quotations should be placed in quotation marks and double-spaced, forming an immediate part of the text. But if a quotation is of a considerable length (more than four or five type written lines) then it should be single-spaced and indented at least half an inch to the right of the normal text margin.
4. **Footnotes:** Footnotes are placed at the bottom of the page on which the reference or quotation, which they identify or supplement, ends. Footnotes should be numbered consecutively and consecutive numbers must be used to correlate the reference in the text with its corresponding note at the bottom of the page, except in the case of statistical tables and other numerical material, where symbols such as the asterisk (*) or the like one may be used to prevent confusion.

Interpretive Phase

Footnotes are always typed in single space though they are divided from one another by double-spaces.

5. **Tables:** Tables should be self-contained and complement but not duplicate, information contained in the text. Tables should be numbered consecutively and double-spaced. Column heading should be brief.
6. **Figures:** A figure is a device that presents statistical data in graphic form. The term figure is applied to a wide variety of graphs, charts, maps, sketches, diagrams and drawings. When skillfully used, figures present aspects of data in a visualized form that may be clearly and easily understood. Figures should not be intended as substitutes for textual description, but included to emphasize certain significant relationships. Figures should be cited in consecutive order in the text.



3.6. Exercise

3.6.1. Tick (✓) the correct answer

1. Typing the research report is done by
 - a. Single space
 - b. Double space
 - c. Multi space
 - d. None of all above.
2. In which section the index of research report is mentioned?
 - a. Summary
 - b. Reference
 - c. Result
 - d. Discussion.

3.6.2. Determine True or False

1. There is no relation between good thinking and effective writing
2. The first page of the report is the title page.
3. Footnotes are placed at the top of the page.
4. The final draft should be written in details.

3.6.3. Fill in the blanks

5. The highlight of the research project is
6. The report should be and objective.
7. Appendices are placed at of research report.
8. Summary section is short in comparison to of the report.

3.6.4. Matching (Match column A with column B)

Column A	Column B
1. The first step in writing report is	i. logical analysis of subject matter
2. Main text provides total outline of	ii. the research report
3. Appendix is written after	iii. bibliography
4. Quotation should be placed in	iv. quotation marks

3.6.5. Short questions

1. What the characteristics of a good research report?
2. What are guidelines for writing a good research report?
3. Discuss the different formatting of research report.
4. Write down the mechanics of research report.