

A CONCISE BOOK OF

NURSING RESEARCH & STATISTICS

(According to Indian Nursing Council Curriculum for B. Sc. Nursing Students)

By

Prof. (Dr.) Sher Singh Morodiya



SINGH PUBLICATION

401, Katewa Nagar, New Sanganer Road,
Jaipur (Rajasthan)

Disclaimer from Authors & Publishers

Knowledge and practice in any field are changing regularly. The new research and experience expand our understanding and changes in research methods. The researchers must always depend on their own experience and knowledge in evaluating and using any information and methods. In using new information, they should be mindful of their own safety and the safety of others, including their professional responsibility.

The publication made every effort to avoid errors or omissions in this book. In spite of this, some errors might have existed in. Any mistake, error or discrepancy noted may please be brought to publication notice which shall be taken care of in the next edition. It is notified that neither the publisher nor the authors or sellers will be responsible for any damage or loss to any one, of any kind, in any manner, there from. The authors and publishers specifically disclaim any implied warranties as to merchantability or fitness of the publication for any particular purpose.

All Rights reserved. No part of this book may be reproduced or copied in any form or by any means (electro-mechanical, photocopying, recording taping or other information storage device, otherwise taping or information retrieval system, etc.), without the written permission of the publisher. Breach of this condition is liable for legal action.

This book has been published in good faith as per the manuscript provided by the author. However the publisher, the printer and the author will not be held responsible for any inadvertent errors(s). All disputes will be subject to jurisdiction of Jaipur under the constitution of India.

Publisher

Published By

Singh Publication

401 Katewa Nagar

New Sanaganer Road

Jaipur 302019

Mail: singhpublicationjaipur@gmail.com

Website: www.singhpublication.in

ISBN: 978-81-959960-0-1

First Edition: 2023

MRP : 350/-

© Publisher

A CONCISE BOOK OF
NURSING RESEARCH & STATISTICS
(According to Indian Nursing Council Curriculum for B. Sc. Nursing Students)

by

Prof. (Dr.) Sher Singh Morodiya

Ph.D., Master of Nursing (RAKCON-DU), MBA (Hospital Management),
MA (English), MA (Distance Education), PG Diploma in Distance Education,
PG Diploma in Higher Education, PG Diploma in Theology,
Bachelor of Journalism & Mass Communication,
Certificate in Computing, Certificate in NABH

Professor & Dean

Faculty of Nursing, Apex University, Jaipur (Rajasthan)

Former

S.No.	Post Held	Place of Posting & Affiliation
1.	Professor & Principal	Mahatma Gandhi College of Nursing & School of Nursing, Sikar (Rajasthan)
2.	Deputy Director (Nursing)	Directorate of Medical & Health Services, Rajasthan, Jaipur
3.	Officer on Special Duty	To the Additional Chief Secretary (Medical & Health), Government of Rajasthan
4.	Nursing Superintendent	SMS Hospital (SMS Medical College), Jaipur (Rajasthan)
5.	Founder Principal	Government College of Nursing, (SP Medical College), Bikaner (Rajasthan)
6.	Faculty	Government College of Nursing, (SMS Medical College), Jaipur (Rajasthan)
7.	Basic Nurse & Nurse Manager/ Nursing in-charge	At various levels of Health Care Setting in Government of Rajasthan

MISS FLORENCE NIGHTINGALE

(Pioneer of Modern Nursing)

12TH MAY, 1820-13TH AUGUST, 1910



THE “NIGHTINGALE PLEDGE”

- ❖ I solemnly pledge myself before God and in the presence of this assembly, to pass my life in purity and to practice my profession faithfully.
- ❖ I will abstain from whatever is deleterious and mischievous and will not take or knowingly administer any harmful drug.
- ❖ I will do all in my power to maintain and elevate the standard of my profession and will hold in confidence all personal matters committed to my keeping and all family affairs coming to my knowledge in the practice of my calling.
- ❖ With loyalty will I endeavour to aid the physician, in his work and devote myself to the welfare of these committed to my care.

Foreword

How do I love thee? Let me count the ways-Elizabeth Barrett Browning



The above sonnet by Elizabeth Barret Browning is the true reflection of many ways we teach and learn the subjects. Every student has a unique way of learning and a teacher has to understand the ways students learn and accordingly change his pedagogy or provide the reading resource in such a way that they can assimilate this information with an ease. In this direction the effort of Prof. (Dr.) Sher Singh Morodiya is remarkable in writing the book "A Concise Book of Nursing Research & Statistics."

Prof. (Dr.) Sher Singh Morodiya asked me to write the forward on his book "A Concise Book of Nursing Research & Statistics" covering the curriculum of B.Sc. Nursing students. I being from the engineering background wondered the reasons for doing this as two disciplines i.e., Engineering and Nursing are poles apart. However, when one is taught research and statistics then somewhere the convergence is taking place and I feel extremely happy to write a forward where these two disciplines have something common to offer. Though the probability and statistics are the topics which require thorough understanding while analyzing data and hypothesis testing but at the undergraduate level they require simple techniques and understanding of basics to begin with. In that perspective the book has done justice to title of the book.

The book addresses the need of research understanding at undergraduate level so that the temperament of analytical behaviour can be inculcated amongst students. In this respect the book is written in a remarkably simple language and easy to understand. All chapters cover all the definition of various terminologies used in research and statistics.

Initial three chapters deal on the research concepts and review of literature which is extremely important to arrive at current state-of-the-art in the subject matter and also to understand the research gap in the area of research. Chapter four and five deal with the sampling and data collection concepts as the entire success of research depends on proper sampling and appropriate data capturing in the process.

Chapter sixth deals with the introduction to statistics and cover basic concepts in statistics and probability with extensive examples which are largely drawn from nursing problems and will go a long way in making in depth understanding of various parameters amongst students of nursing.

Overall the book will be very useful for the students of B.Sc. nursing.

Prof. (Dr.) O.P. Chhangani

Vice Chancellor, Apex University, Jaipur

Ex-Pro. Vice Chancellor, Rajasthan Technology, Kota

Contributors of the Book



Prof. (Mrs.) Cimil Babu

Ph.D. Scholar, M.N.(OBG-RAKCON-DU)
Professor, Holy Family College of Nursing
(D.U.), Okhla Road, New Delhi



Air Commodore **Mr. Devendra Singh Shekhawat**, VSM (Retired) M.Phil
(Defence & Strategic Std.) , M. Tech (IIT Kgp)
Dean, Academics, Apex University, Jaipur



Prof. (Dr.) Narendra Kumar Kaushik

Ph.D. M.N.(MSN-RAKCON-DU)
Principal, Government College of
Nursing, Bikaner, (Rajasthan)



Mr. Roop Singh

M.A., M.Ed.
Principal (Retired)
Rajasthan School Education



Mr. Ritu Parn, M.Tech.

Director
3S Industries Pvt. Ltd,
Alwar (Rajasthan)



Ms. Ekta Phansal

B.Sc., B.Ed.
Certificate in Fashion Designing (INIFD,
Pune) Pursuing - M.Sc (Botany)

Reviewers of the Book



Dr. Binu Joe, Rajkot

Founder President, Nursing Scholar
Society, Principal, Christ College of
Nursing, Rajkot (Gujrat)



Dr. B.L. Sharma

Ph.D., M.Sc. (MSN)
Lecturer, Government College of
Nursing, Jaipur (Rajasthan)



Mrs. Deby Chakraborty

Ph.D. Scholar, M.N.(OBG-RAKCON-DU)
Assistant Professor, Govt. College of
Nursing, Agartalla (Tripura)



Mrs. Gulshan Roy Chowdhury,

Ph.D. Scholar, M.N.(OBG-RAKCON-DU)
Lecturer, College of Nursing
ABVIMS & Dr. R.M.L. Hospital, New Delhi



Mr. Hemant Saini

M.Sc. (Nephro-Urology Nursing-AIIMS)
Associate Professor, Baba Farid University
of Health Sciences, Faridkot (Punjab)



Mr. Hemant Kumar

M.Sc. (MSN), MA (Psychology), DSPR
Assistant Professor, College of Nursing, Govt.
MRA Medical College, Ambedkarnagar (UP)



Dr. Jagdeep Saini

Ph.D., M.N.(CHN-RAKCON-DU), MBA
(HA), Senior Faculty, Government
College of Nursing (SMSMC), Jaipur



Prof. Jitendra Chauhan

M.Sc. (Child Health Nursing) , Principal
Government College of Nursing
Jamnagar (Gujrat)



Mr. Kailash Chandra Saini

Ph.D. Scholar, M.Sc.(MSN), Associate
Professor, Apex School of Nursing,
Apex University, Jaipur (Rajasthan)



Mrs. Lalremuati

M.N.(CHN-RAKCON-DU)
Tutor, College of Nursing, RIPANS
(An autonomous body under Ministry of
Health and Family Welfare) Aizawl (Mizoram)



Dr. Madan Mohan Meena

Ph.D., M.Sc. (CHN), MA (PA), MBA (HM)
Principal, Rajasthan University of Health
Sciences, College of Nursing, Jaipur



Dr. Rahul Tiwari

Principal, Ravi College of Nursing
EH-2, Block- D, Shastripuram
(Near EPIP Centre) Agra-282007(UP)



Professor (Dr.) K. Rajarani

Ph.D. M.Sc. (MSN), Professor
College of Nursing, Govt. MRA Medical
College, Ambedkarnagar (UP)



Mr. Upendra Yadav

M.N.(Child Health Nursing-RAKCON-DU)
College of Nursing, B.P. Koirala Institute
of Health Sciences, Dharan, Nepal

Preface

The prescribed course of nursing research and statistics is designed to enable undergraduate students to develop an understanding of basic concepts of research studies in various settings and utilize the research findings to provide quality nursing care. There are so many books available on the subject in market and they have materials more than need. The learner might confuse to select the needed material for their learning. During my teaching of this subject, I felt the need of such book, which contains learning material according to their prescribed syllabus. It is an effort to fulfill the gap. This book is written in simple English language so that the students may understand the content easily. It is an outgrowth of my experience of teaching and research.

Regarding to organization of book, it contains six chapters. Chapter one is related to research and research process. Chapter two presents review of literature. Chapter three tells about research approaches and research designs. Chapter four presents the details of sampling and data collection. Chapter five explains analysis, interpretation, communication and utilization of data. Chapter six is an introduction to statistics.

At the starting of each chapter, there is given a “structure” of the chapter. This “structure” of the chapter includes the names of the main heading and sub-headings. In the last of each chapter, summary of the chapter is given under heading “let us sum up.” After completion of each or some heading, self-evaluation exercises are given under heading “let us check our progress.” It helps learners/students to evaluate their own learning and help to make confident in that subject areas. In the last of the chapter, answers of heading “let us check our progress” are given under heading “answers to let us check our progress.” It helps learners/students to compare their answer with the actual answers. This book is written in a simple and understandable English language, so students/learners can learn without help of teacher too. It is written in “we” form in spite of “you” form; it means the author is also a learner with all. In the sixth chapter, introduction to statistics, each topic is written in simple and understandable language. Each topic contains one solved statistical exercise and three unsolved exercises under “let us check our progress.” At the end of sixth chapter; solutions/calculations of these three exercises are given under “answers to let us check our progress.”

This book is primarily intended to serve as a textbook for undergraduate students of nursing discipline. But it may also helpful for those teachers who are teaching the nursing research and statistics to nursing students.

I feel amply rewarded if this book proves user-friendly at the hands of readers/learners.

Every effort has been made to make the book error free but nothing can be perfect and some errors may still remain in the book. I shall feel obliged if any such errors are brought to my notice. I look forward for feedback/suggestions from readers and experienced teachers for further development and improvement of the book as a whole.

I sincerely invite the feedback/suggestions to singhpublicationjaipur@gmail.com

Prof. (Dr.) Sher Singh Morodiya

Acknowledgements

First of all, I would like to express my deep sense of gratitude to the omnipresent, merciful, fountainhead of all consciousness in the cosmos; Ra-Dha-Sva-Aah-Mi Dayal for grace which made it all possible.

I would like to express my deep heartfelt gratitude to my parents who taught me to walk on the path of hard work, truth and success.

I wish to acknowledge my gratitude to all my teachers, mentors, guides, advisors, counselors and seniors who have contributed to my success in my career.

I would like to express my sincere gratitude to Hon'ble Prof. (Dr.) O.P. Chhangani, Vice Chancellor, Apex University, Jaipur, who accepted my request and obliged me by writing the "Foreword" for this book.

I express my gratefulness to my beloved wife, Resham for her light-hearted and positive attitude and also her continuous encouragement and cooperation to introduce this book.

I am grateful to all those persons whose writings have helped me in writing of this book. I am equally grateful to the contributors as well as the reviewers of the manuscript of this book who made extremely valuable suggestions and has thus contributed in enhancing the level of the book.

I would like to thank the patience, love and support that I received from Nisha (my daughter-in-law), Drishti (my grand daughter) and all my family members.

I would like to place on record the appreciation to Singh Publication, Jaipur for their untiring efforts and cooperation in bringing out the book in time and in such an elegant form.

I am indeed thankful to Mr. Shyam Shonkiya, page-setter for changing my dream into reality, without their help it would be a dream only.

Prof. (Dr.) Sher Singh Morodiya



From the Publication

- ❖ It gives us immense pleasure to share with you that "Singh Publication" is a new initiative in field of publication.
- ❖ Keeping in mind, the New Education Policy of our country and changing trends in nursing education, "Singh publication" have taken up a mission of such books, those will student-friendly, syllabus based, written by subject experts in whole country.
- ❖ In this light, the Singh Publication aimed to those students who are unable to understand the books available in difficult English language.
- ❖ To fulfill this gap, it has been published first easily understandable book, named, "A concise Book of Nursing Research & Statistics."
- ❖ The authour of this book as well as initiator of the publication is one person who has lot of experience in areas of clinical, teaching, guiding, research, writing, editing, translating and administration/ management in government as well as private sector.
- ❖ The growth of the publication cannot possible without invaluable contribution of reader, authour, reviewer, contributor and recommender, these all are pillars of any publication.
- ❖ We need your support and contributions. We will happy and grateful to you if you join hands with us in form of author, contributor and reviewer.
- ❖ Let's us join together and share ideas and knowledge in field of your expertness.
- ❖ We are looking forward to your cooperation in future as well. Share your CV to singhpublicationjaipur@gmail.com

About the Authour

Prof. (Dr.) Sher Singh Morodiya completed Diploma in General Nursing in 1983 from Government School of Nursing, S.K. Hospital, Sikar (Rajasthan). He got his B.Sc. Nursing Degree in 2004 from IGNOU, New Delhi. He received his Master of Nursing Degree (Medical Surgical Nursing) in 2008 from Raj Kumari Amrit Kaur College of Nursing, University of Delhi. He has been awarded Ph.D. Degree in 2013 from Pacific Academy of Higher Education and Research University, Udaipur, Rajasthan.



Apart from the Nursing field, he completed Bachelor's Degree in Arts and Master's Degree in Arts (English) from University of Rajasthan. He got MBA Degree in Hospital Management from Algappa University, Tamilnadu. He completed Master's Degree in Arts (Distance Education) from IGNOU. He also did Post-Graduation Diploma in Distance Education and Post-Graduation Diploma in Higher Education from IGNOU. He was awarded Post Graduation Diploma in Theology from Dayalbagh Educational Institute, Agra (U.P.) and Bachelor's Degree in Journalism and Mass Communication from Vardhman Mahaveer Open University, Kota (Rajasthan).

Experience: He started his career after General Nursing Diploma as Nurse Grade II (Basic Nurse) at Government dispensary in 1984. He worked at various health care settings i.e. Primary Health Centre, Community Health Centre, District Hospital and Medical College Hospital in various capacities as Basic nurse, Nurse Manager/ Ward In-charge, Nursing Superintendent. He experienced as Lecturer and Founder Principal, Government Colleges of Nursing and engaged in Under Graduate and Post Graduate programmes. He got chance to work on the post of the Officer on Special Duty to the Additional Chief Secretary (Medical & Health) along with the responsibility of Deputy Director (Nursing) i.e. highest post in Government of Rajasthan for nursing personnel.

He is having **39 Years of experience in the field of Nursing**. He has lot of confidence in area of teaching, guiding, examining/evaluating, care of clients, research and management/administration at various levels of health care settings. Now he is working as Professor, Principal & Dean, Faculty of Nursing, Apex university, Jaipur. Here, he is also registered as research supervisor/guide for Doctoral Programmes.

He is authour of one book named; "Performance Appraisal of Nursing Personnel (Issues and Challenges)." He has been translated six nursing books from English to Hindi. He is Editor-in-Chief of International Journal of Medical Surgical Nursing and member of editorial board in seven peer reviewed/refereed journals from India and two foreign (USA) journals. He published 26 articles in International & National Journals, one Article Published in Conference Proceeding Book, eight articles published in News Papers.

He has been presented seven research papers in International Conferences and four in National conferences. He has received **BEST PAPER AWARD** in an International Conference held at Pacific University, Udaipur in 2019. He is life time member in ten professional bodies.

Syllabus for B.Sc. Nursing

Placement- Fourth Year
Internship

Time: Theory- 45 Hours
Practicals-45 Hours

Course Description: The course is designed to enable students to develop an understanding of basic concepts of research, research process and statistics. It is further, structured to conduct/ participate in need based research studies in various settings and utilize the research findings to provide quality nursing care. The hours for practical will be utilized for conducting individual/group research project.

Unit	Time (Hrs)	Learning Objectives	Contents	Teaching Learning Activities	Assessment Methods
I	4	Describe the concept of research, terms, need and areas of research in nursing. Explain the steps of research process.	Research and research process Introduction and need for nursing research Definition of research & nursing research. Steps of scientific methods. Characteristics of good research. Steps of research process-overview.	Lecture, discussion. Narrate steps of research process followed from examples of published studies.	Short answer. Objective type.
II	3	Identify/state the research problem and objectives	Research problem/question Identification of problem area. Problem statement. Criteria of a good research problem. Writing objectives.	Lecture, discussion. Exercise on writing statement of problem and objectives.	Short answer. Objective type.
III	3	Review the related literature	Review of literature Location Sources On line search CINHAL, COCHRANE etc. Purposes Method of review	Lecture, discussion. Exercise on reviewing one research report/article for a selected research problem. Prepare annotated-bibliography.	Short answer. Objective type.

Unit	Time (Hrs)	Learning Objectives	Contents	Teaching Learning Activities	Assessment Methods
IV	4	Describe the research approaches and designs.	Research approaches and designs. Historical, survey and experimental Qualitative and quantitative designs.	Lecture, discussion. Explain types of research, approaches used from examples of published and unpublished research studies with rationale.	Short answer. Objective type.
V	8	Explain the sampling process. Describe the methods of data collection.	Sampling and data collection. Definition of population, sample, sampling criteria, factors influencing sampling techniques. Data- why, what, from whom, when and where to collect. Data collection methods and instruments. Methods of data collection. Questioning, interviewing, observation, record analysis and measurement. Types of instruments Validity & reliability of the instrument. Pilot study. Data collection procedure	Lecture, discussion. Reading assignment on examples of data collection tools. Presentation of sample data collection tool. Conduct group research project.	Short answer. Objective type.
VI	4	Analyze, interpret and summarize the research data Analysis of data	Analysis of data Compilation, tabulation, classification, summarization, presentation, interpretation of data.	Lecture, discussion. Preparation of sample tables.	Short answer. Objective type.

Unit	Time (Hrs)	Learning Objectives	Contents	Teaching Learning Activities	Assessment Methods
VII	15	Explain the use of statistics, scales of measurement and graphical presentation of data. Describe the measures of central tendency, variability and methods of correlation.	Introduction of Statistics Definition, use of statistics, scale of measurement. Frequency distribution and graphical presentation of data. Mean, median, mode, standard deviation. Normal probability and tests of significance. Co-efficient of correlation. Statistical packages and its application.	Lecture, discussion. Practice on graphical presentation. Practice on computation of measures of central tendency, variability & correlation.	Short answer. Objective type.
VIII	4	Communicate and utilize the research findings.	Communication and utilization of research. Communication of research findings-verbal reports, writing research report, writing scientific article/ paper, critical review of published research. Utilization of research findings.	Lecture, discussion. Read/presentation of a sample published/unpublished research report. Writing group research project.	Short answer. Objective type. Oral presentation. Assessment of group research project.

CONTENTS

Contents		Pages
Foreword		v
Contributors & Reviewers of the Book		vi
Preface		vii
Acknowledgements		viii
From the Publication		ix
About the Authour		x
Syllabus of Nursing Research and Statistics		xi
Chapter-1	Research and Research Process	1-16
	Research and Research Process	
	1. Objectives	2
	2. Introduction	2
	3. Basic terms used in research	2
	4. Research problem/ question 4.1 Research problem 4.2 Research question 4.3 Identification of problem area or sources 4.4 Criteria of a good research problem	4
	5. Research and nursing research 5.1 Definition of research 5.2 Definition of nursing research 5.3 Characteristics of good research 5.4 Importance of nursing research	5
	6. Formulation of research objectives 6.1 Meaning 6.2 Need 6.3 Characteristics 6.4 Types	7

	6.5 Methods	
	7. Purpose of research 7.1 Description 7.2 Exploration 7.3 Prediction	8
	8. Scientific methods 8.1 Basic postulates 8.2 Purposes 8.3 Steps	9
	9. Research process	10
	10. Stages of research process 10.1 Planning 10.2 Implementation of research plan 10.3 Communicating and applying result	11
	11. Steps of research process 11.1 Statement of research problem 11.2 Defining purposes/ objectives 11.3 Defining terms 11.4 Review of literature 11.5 Identifying assumptions 11.6 Operational definition 11.7 Stating hypotheses 11.8 Selecting research design 11.9 Planning for data collection 11.10 Data collection, analyzing and interpretation of results 11.11 Writing the report, critiquing, publication and application of results	11
	12. Let us sum up	15
	13. Answers to check our progress	15
Chapter-2	Review of Literature	17-23
	1. Introduction	17
	2. Objectives	17

	3. Review of literature 3.1 Meaning 3.2 Purposes 3.3 Types 3.4 Source 3.5 Method/steps of review	17
	4. Let us sum up	22
	5. Answers to let us check our progress	23
Chapter-3	Research Approaches and Designs	24-33
	1. Introduction	24
	2. Objectives	24
	3. Research approaches 3.1 Meaning	24
	4. Types of research 4.1 Historical research 4.2 Survey research 4.3 Experimental research	24
	5. Research design 5.1 Meaning 5.2 Types -qualitative and quantitative	28
	6. Let us sum up	23
	7. Answers to check our progress	33
Chapter-4	Sampling and Data Collection	34-69
	1. Introduction	34
	2. Objectives	34
	3. Sampling 3.1 Concepts 3.2 Purposes or need 3.3 Advantages 3.4 Disadvantages 3.5 Characteristics of a good sample	35

	3.6 Suitability 3.7 Sampling process 3.8 Types of sampling techniques/ methods	
	4. Data collection 4.1 Concept 4.2 Techniques 4.3 Tools 4.4 Pilot study 4.5 Data collection procedures	47
	5. Data organization	66
	6. Data presentation	66
	7. Let us sum up	67
	8. Answers to let us check our progress	68
Chapter-5	Analysis, Interpretation, Communication and Utilization of Data	70-110
	1. Introduction	70
	2. Objectives	71
	3. Concepts used in analysis 3.1 Tabulation 3.2 Compilation 3.3 Summarization	71
	4. Analysis of data 4.1 Definition 4.2 Purpose 4.3 Types or classification	71
	5. Interpretation of data 5.1 Meaning 5.2 Techniques 5.3 The points may consider in interpretation	72
	6. Communication of research findings 6.1 Writing research reports on research findings 6.2 Verbal or oral presentation of research reports 6.3 Poster presentation	73

	6.4 Writing scientific article/ paper	
	7. Utilization of research findings	100
	8. Critical review or evaluation of published research	103
	9. Let us sum up	109
	10. Answers to check our progress	110
Chapter-6	Introduction to Statistics	112-196
	1. Introduction	112
	2. Objectives	112
	3. Statistics: 3.1 Basic terms 3.2 Definition 3.3 Characteristics 3.4 Limitations 3.5 Use of statistics 3.6 Importance of statistics 3.7 Types of statistics	112
	4. Measurements in statistics 4.1 Data 4.2 Continuous and discrete Series 4.3 Scale/levels of measurement 4.4 Determining the meaning of test score in a continuous series	120
	5. Descriptive statistics 5.1 The Frequency distribution 5.2 Graphic presentation of the frequency distribution 5.3 Measures of central tendency 5.4 Measures of variability	123
	6. Normal probability	146
	7. Test of significance 7.1 Meaning 7.2 P-Value 7.3 The significance levels 7.4 The level of confidence	147

	7.5 Types of errors	
	8. Co-efficient of correlation 8.1 Meaning 8.2 Types and degree of co-relation 8.3 Methods of determining co-relation	148
	9. Statistical packages and its application 9.1 Statistical packages meaning 9.2 Application of statistical packages	159
	10. Let us sum up	161
	11. Answers to check our progress	162
	Bibliography	198

CHAPTER-1

Research and Research Process

Structure

1. Introduction
2. Objectives
3. Basic terms used in research
4. Research problem/ question
 - 4.1 Research problem
 - 4.2 Research question
 - 4.3 Identification of problem area
 - 4.4 Criteria of a good research problem
5. Research and nursing research
 - 5.1 Definition of research
 - 5.2 Definition of nursing research
 - 5.3 Characteristics of a good research
 - 5.4 Importance of nursing research
6. Formulation of research objectives
 - 6.1 Meaning
 - 6.2 Need
 - 6.3 Characteristics
 - 6.4 Types
 - 6.5 Methods
7. Purpose of research
 - 7.1 Description
 - 7.2 Exploration
 - 7.3 Prediction
8. Scientific methods
 - 8.1 Basic postulates
 - 8.2 Purposes
 - 8.3 Steps
9. Research process
10. Stages of research process
 - 10.1 Planning
 - 10.2 Implementation of research plan
 - 10.3 Communicating and applying result
11. Steps of research process
 - 11.1 Statement of research problem
 - 11.2 Defining objectives
 - 11.3 Defining terms
 - 11.4 Review of literature

- 11.5 Identifying assumptions
- 11.6 Operational definition
- 11.7 Stating hypotheses
- 11.8 Selecting research approach
- 11.9 Planning for data collection
- 11.10 Data collection, analyzing and interpretation of results
- 11.11 Writing the report, critiquing, publication and application of results
- 12. Let us sum up
- 13. Answers to check our progress

1. Introduction

We can gain knowledge through various methods and media like books, journals, newspapers, magazines, reports, records, seminars, conferences, television, radio, social media etc. Research is also a method of gaining knowledge. Research is a systemic process of investigation to find out new knowledge. Research travels through an established stages and steps. It is called research process.

In this chapter, we will define research and nursing research and also explain the steps of scientific methods. We will describe the characteristics of a good research. We will know the importance and purposes of nursing research. We will explain the stages and the steps of research process.

2. Objectives

After studying this chapter, we will be able to:

- 1. Understand basic terms used in research
- 2. Define research and nursing research.
- 3. Describe the characteristics of good research.
- 4. Know the importance and purpose of nursing research.
- 5. Explain the steps of scientific methods.
- 6. Explain the stages of research process.
- 7. Describe the steps of research process.

3. Basic terms used in research

Following are the basic terms in research:

Research abstract- An abstract is a brief description of a complete research. It is description of the work in concise, self-contained and clear. It is a short piece of writing that tells the main contents of a research. In research journals, it is written at the beginning of an article.

Research data- Data is a plural number, its singular number is datum. It is facts, figures, evidence, general material or knowledge. It is collected, observed, generated or created during the process of research study.

Variable- Variable is a situation, number or quantity, which can vary. It is a condition that is not staying the same, it is often changing.

Dependent variables- Dependent variables are those that depend on other factors. As soon as the independent variables are manipulated, they change accordingly. These variables are also called the criterion variables.

Independent variables- Independent variables are those that are not depend on other factors, so they do not change. The independent variables are purposefully manipulated. These variables are also called the manipulated variables.

Research variables- Research variables are the characteristics, traits, attributes or qualities. These are measured or observed in a natural setting.

Demographic variables- These are independent variables because they cannot be manipulated. Characteristics, traits, attributes or qualities of study population are called demographic variables, e.g. age, gender, religion, educational qualification, professional qualification, professional experience, marital status, occupation, income, etc.

Extraneous variables- These are the factors that are not the part of research study. They may affect the study variables. If left uncontrolled, they may lead to inaccurate results.

Operational definition- It allows the researcher to describe in a specific way, they mean when they use a certain term. The researcher clarifies and defines terms that how he/she will be observed and measured in the study.

Conceptual framework- It includes a less formal and less developed mechanism in comparison to theories. It is an analytical tool with several variations and contexts. It is composed of one or more formal theories. It has concepts and empirical findings from the literature. It is used to show relationships among these ideas and their relationship to the research study.

Assumptions- It is a statement that is believed to be true. It has not been scientifically tested. These are principles; those are accepted without proof or verification only logically explained.

Hypothesis- It is a statement of expectation or prediction that can be tested by research. It is tentative result of research. It is educated or calculated guess by researcher proposes to test or enquiry.

Review of literature- It is a survey of available literature related to a specific chosen research topic or a research question, e.g. books, journals, theses etc. It is done to identify gaps.

Delimitation- There are many forms of difficulties and they required solution. But due to lack of time, facilities, finance etc. are not possible, so prior to study, these are considered. These are decided during planning stage that beyond some cut off points, the researcher does not want to go.

Limitations- Limitations are the weakness, shortcomings, conditions or influences of the entire study, perceived by the researcher. These are not controlled by the researcher. These are experienced during implementation stage and should report while writing.

Manipulation- It is skillful handling or controlling of something or someone. It is a process of initiating intervention by the researcher in experimental or quasi-experimental study. It

is done to see the effect of the independent variable on dependent variables.

Population- Population is a set or a group of individuals or objects that possess some traits or characteristics in which, the researcher has some interest.

Target population- A target population is a set or a group of individuals or objects that researcher wants to generalize the study findings.

Accessible population- Accessible population is a set or a group of individuals or objects that actually available for the study by the researcher.

Research study settings- It is the area where the research is conducted. The chosen area could be natural, partially or fully controlled.

Sample- A sample is a part or an amount of the population to represent the entire population. It may be a group of items or objects or people.

Sampling- The process of selecting a sample or a part or an amount of the population to represent the entire population of the study.

Probability sampling- In this type of sampling, the probability is ensured. Each element has an equal chance of being selected. These are randomly selected.

Non probability sampling- In this type of sampling, the probability is not ensured. Each element has not an equal chance of being selected. These are not selected randomly.

Reliability- It is the degree to which a measuring tool/instrument gives the same results each time that it is used.

Validity- It is the degree to which the tool/instrument measures, what it is intended to measure.

Pilot study- The pilot study is done to find out the feasibility of the study. It is used to test the tool and select the appropriate methodology. It also identifies the potential problems in research study.

Analysis- It is categorizing, ordering and summarizing the research data statistically. It is done to obtain the answers to research questions.

Interpretation- It is the study of results of the analysis in order to make inferences about the occurrence of relationship.

4. Research problem/ question

4.1 Research problem

Research problem is that which needs a researcher to get the best answer or solution for the selected problem. It is a mysterious condition about a related field, some difficulty in theory or practice that reveals the requirement of investigation.

4.2 Research question

Research question is a statement of specific query to which, researcher wants to search an answer of the question.

4.3 Identification of problem area or sources

The following sources or problem areas may be identified by any researcher:

- (a) An investigation of problems which are being faced by nursing teachers, nursing students and their guardians may tell many topics which may be investi-

gated by researcher.

- (b) Text books and research journals are good sources of problems. Many research articles suggest problems for further research. They also demonstrate use of many techniques in research in a certain problem.
- (c) Needy persons may contact senior researchers, research academicians or advisers who are considered most competent persons for this purpose.
- (d) Changes and innovations in field of health profession bring new problems which provide new opportunities for new investigators.

4.4 Criteria of a good research problem

The criteria of a good research problem is as below:

- (a) The problem should be significant. There should be enough variables for investigations. The answer or solution of a research of problem should make an important difference in current theories or practice.
- (b) The problem should be solved by the process of research.
- (c) The problem should be new and original.
- (d) The problem should have a theoretical value or to fill the gap in the literature.
- (e) The problem should be workable. A researcher should competent to carry out the research on the selected topic. He should also have enough time, adequate financial resources and have cooperation of the population on the identified research problem.
- (f) There should have availability of good tools for data collection.

Let us check our progress-1

- (i) *Write the meaning of research question.*
- (ii) *State the criteria of a good research problem.*

5. Research and nursing research

5.1 Definition of research

Research: The word “research” is derived from the French term *recherche*, composed of a prefix, *re* and a verb, *search*. *Re* means "once again" and *search* means "to examine closely and carefully." It means, research is a close and careful examination of the facts and their relationship to discover new knowledge.

“Research is a systematized effort to gain new knowledge.” - *Redman and Morry, 1923*

“Research is the manipulation of things, concepts or symbols for the purpose of generalizing to extend, correct or verify knowledge, whether that knowledge aids in construction of theory or in the practice of an art.” - *D. Slesinger and M. Stephenson*

“A careful investigation or inquiry specially through search for new facts in any branch of knowledge.” - *The Advanced Learner's Dictionary of Current English, Oxford, 1969*

“Research may be defined as the systematic and objective analysis and recording of controlled observation that may lead to the development of generalizations, principles, theories, resulting in prediction and possible ultimate control of

events.” - *J. W. Best, 1969*

“Research is the process of systematically obtaining accurate answers to significant and pertinent questions by the use of the scientific method of gathering and interpreting information.” - *Clover and Balsley, 1969*

“Research may be defined as planned, systematic search for information for the purpose of increasing the total body of man's knowledge. It involves looking for information which at the time is not available or for which that has no generally accepted evidence.” - *Archold Lancaster, 1982*

“Scientific research is a systematic, controlled, empirical and critical investigation of natural phenomena guided by theories and hypotheses about the presumed relations among such phenomena.” - *Fred N. Kerlinger, 1986*

5.2 Definition of nursing research

- (i) “Nursing research refers to the use of systematic, controlled, empirical and critical investigation in attempting to discover or confirm facts that relate to specific problem or question about the practice of nursing.” - *Walls and Bausell, 1981*
- (ii) “Nursing research develops knowledge about health and promotion of health over the full lifespan, care of person with health problems and disabilities to respond effectively to actual or potential health problems.” - *Commission of Nursing Research, American Nurses Association, 1981*
- (iii) “Nursing research is a way to identify new knowledge, improve professional education and practices and use of resources effectively.” - *International Council Nurses, 1986*
- (iv) “Nursing research is defined as the systematic detailed attempt to discover or confirm facts that relate to a specific problem to improve the practice and profession of nursing.” - *Abdellah and Levine, 1994*
- (v) “Nursing research is defined as systematic search for knowledge about issues of importance to nursing.” - *Polit and Hungler, 2001*
- (vi) “Nursing research is a scientific process that validates and refines existing knowledge and generates new knowledge, which directly and indirectly influences nursing practice.” - *Burns and Grove, 2005*
- (vii) “Nursing research is a systemic approach to gather information for the purpose of answering questions and solving problems in the pursuit of creating new knowledge about nursing practice, education and policy.” - *Hek and Moule, 2006*

Let us check our progress-2

- (i) Define research.
- (ii) Define nursing research

5.3 Characteristics of a good research

- (i) **Systemic:** It means, structured; it has steps in a sequence. The sequence exists in the well-defined set of rules. It does not follow creative thinking. It rejects the use of guessing and intuition in reaching at conclusion.

- (ii) **Logical:** Logical means research is guided by rules of logical reasoning. Process of induction and deduction are also pay value in carrying out research. Logical reasoning makes research more meaningful.
- (iii) **Empirical:** It is a process in which the information or data are observed through our sense organs. The gained knowledge is based on objective evidences.
- (iv) **Valid and verifiable:** It means our findings are correct and can be verified by us and others.
- (v) **Replicable:** It means the results of research can be verified by replicating the study and building a sound basis for decision making.
- (vi) **Generalization:** It is a characteristic of a good research. It says that can we go beyond the study situation or population? It is drawing broad inferences from particular observations.

5.4 Importance of nursing research

- (i) **Need for development of nursing knowledge:** Like other professions, nursing profession also depends on specialized and verified knowledge and skills. The nursing profession has a social responsibility toward society. So, it is essential to contribute practice with advanced knowledge and skills. There is need to develop theories for nursing practice. It would add body of knowledge and skills. It may be shared by other disciplines.
- (ii) **Research to improve practice of nursing:** There is new tendency to build nursing theories. Therefore, nursing researchers has a responsibility of developing a scientific basis for nursing practice.
- (iii) **Research to improve health care delivery:** In our country, present scenario of health care delivery system is imitative and intuitive. It should innovative and creative. It should also based on scientific research. There is need of research on quality of providing care. Research findings enable the health care delivery system to provide more efficient as well as cost-effective care. There are many problems for researchers in health practice. These are concerned with planning, management, logistics and delivery of health care services.

6. Formulation of research objectives

6.1 Meaning

Research is a careful and detail study of a selected problem with the help of scientific method. Research objectives are the outcomes that we aim to achieve by conducting research. The clearly defined/written objectives are very important to get an answer or a solution of a problem. These are developed/framed/formulated in planning stage of research.

6.2 Need

The following are the needs that will help the researcher:

- (i) **Focus-** The clearly defined/written objectives are helpful in focusing on the research/study. These helps in concentration on ultimate goal. It prevents the

wastage of resources (man, money and material).

- (ii) **Avoid-** The clearly developed/framed/formulated objectives are helpful in avoiding the collection of unnecessary/unwanted data.
- (iii) **Organize-** The clearly defined/written objectives help in organizing the phases of research.
- (iv) **Directions-** The clearly defined/written objectives provide a right direction to develop a research methodology. It will guide in collection, analysis, interpretation and utilization of data.

6.3 Characteristics

Research objectives should be SMART i.e. specific, measurable, attainable, realistic and time bound.

1. It should be relevant, feasible, logical, observable, unequivocal and measurable.
2. It should be achieved within the resources and time bound.
3. It should obtain answer of research question.
4. It should test research hypothesis.

6.4 Types

There are two types of research objectives

- (i) **General objectives-** These state what we expect to achieve in general terms. These are usually less in number. These are broad goals for achievement.
- (ii) **Specific objectives-** These are made by breaking down the general objectives into smaller. These are logically connected parts that say the various aspects of problem.

6.5 Methods

The following points should keep in mind while stating the research objectives

- (i) The research objectives should not more than 5-6 in numbers. More than 5-6 objectives may overload and the results of study may vague and ambiguous.
- (ii) These are clearly framed in operational terms that what the researcher is going to do, where to do and for why to do.
- (iii) These should be clear, concise, attainable, quantifiable and measurable.
- (iv) These should be formulated with the use of action verb, so it may be evaluated easily.

7. Purpose of research

There are several purposes of research. The main purpose of research is to discover or find out answer to questions through research process or scientific procedures. By this process, the truth is discovered yet which is hidden. Though, each study has its own purpose, which depends on the nature of research study. There are three specific purposes, which are explained below:

- 7.1 Description:** Its main purpose is to describe of the state of affairs as it exists at present. In this type of purpose, the researcher has no control over the variables. He can only report, what is happening. This purpose or method is utilized in all

kinds of descriptive research as survey methods. The main objective is to describe systematically a situation or area of interest factually and accurately. It helps in identifying the factors present in the problem.

7.2 Exploration: Its main purpose is to develop the hypotheses, not testing. Such types of research includes case-study methods/ in depth approaches to reach the basic casual relations. It focuses on relationship between phenomena or the cause and effect relationship.

7.3 Prediction: It is mainly concerned with forecasting outcomes, consequences or effects. It includes the situations where the research has ability to predict. It also include to some extent can explain the occurrence of specific phenomenon. These findings would contribute to health care practices.

Let us check our progress-3

- (i) Write the characteristics of research objectives.
- (ii) Describe purposes of research.

8. Scientific methods

The scientific method is an objective, logical and systemic. It aims to build general knowledge about natural phenomena. It is free from personal bias or prejudice. It is a method to ascertain demonstrable qualities of a phenomenon capable of being verified. In this method, the researcher follows the rules of logical reasoning in an orderly manner.

8.1 Basic postulates or axioms

- (i) Scientific method relies on empirical evidences.
- (ii) It follows the systematic and orderly processes.
- (iii) It follows the objective considerations.
- (iv) It utilizes relevant concepts.
- (v) It follows ethical neutrality.
- (vi) It results into probabilistic predictions.
- (vii) It is based on assumptions or hypotheses.
- (viii) It is used to develop or test theories.

8.2 Purposes

Its purposes are description, exploration, prediction, explanation, control and identification of relationship of the facts.

8.3 Steps

- (i) Identifying and defining the problem.
- (ii) Defining the objectives of the studying.
- (iii) Reviewing the literature related to the problem-solving.
- (iv) Defining the objectives of the studying.
- (v) Identifying variables and assumption.
- (vi) Forming hypotheses.
- (vii) Determining ethical considerations.
- (viii) Describing research design.

- (ix) Defining the population, sample and sampling techniques.
- (x) Planning for data collection.
- (xi) Planning for analysis and interpretation.
- (xii) Collecting data.
- (xiii) Analysis and interpretation of data.
- (xiv) Communicating the findings of the study.

Let us check our progress-4

Enumerate the steps of scientific methods.

9. Research process

The Process means “a course of action or proceeding.” The research process is a process of multiple scientific steps in conducting the research work. Each step is interlinked with other steps. Here, the research process is described in three stages. These stages are:

1. Planning
2. Implementing
3. Communicating and applying results.

Stages	Steps
1. Planning	1. Conceptualization of research problem <ul style="list-style-type: none"> (i) Statement of research problem (ii) Defining purposes/objectives (iii) Defining terms 2. Review of literature 3. Development of conceptual framework <ul style="list-style-type: none"> (i) Identifying assumptions (ii) Defining variables (iii) Stating hypotheses 4. Research design <ul style="list-style-type: none"> (i) Selecting research approach (ii) Planning for data processing (iii) Planning for data analysis (iv) Planning for interpretation
2. Implementing research plan	1. Data collection
	2. Data analysis
	3. Interpretation of results
3. Communicating and applying results	1. Writing the report
	2. Critiquing and publication
	3. Application of results

These stages are further divided in more steps.

10. Stages of research process

10.1 Planning

The planning stage of research process includes the following interdependent steps which include:

1. Conceptualization of research problem
2. Review of literature
3. Development of a conceptual frameworks
4. The research designs

Many of these steps are overlapping in same time, e.g. review of literature is usually carried out during above step number 1, i.e. (conceptualization of research problem), 3, i.e. (development of a conceptual frameworks) and 4, i.e. (the research design).

Conceptualization of research problem may include statement of research problem, defining purposes and defining terms.

Development of conceptual framework includes identifying assumptions, define variables and stating hypotheses.

Research design consist the steps of selecting research approach, planning for data processing, planning for data analysis and planning for interpretation.

Pilot study use to test the feasibility of the design.

10.2 Implementation of research plan

Implementation stage comprises data collection, data analysis and interpretation of results. The researcher collects the data with the help of reliable and valid tools, further he/she analyses the collected data through descriptive and inferential statistics. He/she also interpret his findings in various forms of presentation but he/she analyses the data objectively, so his/her opinion or view is not presented.

10.3 Communicating and applying results

Communicating and applying results are the last stage of the research process. It involves writing the report, critiquing, publication and application of results. The reports are written in different forms for different interest group members i.e. thesis or dissertation for fulfilling an academic requirement of university and an article in professional journal or newspaper etc. The forms may different; but it always includes the research design, method of data collection, analysis and conclusion of the study. It also includes the implications of findings in field of education, service and care.

11. Steps of research process

11.1 Statement of research problem

It is the first step of the research process. Statement of research problem means to define a problem with specification in detail with precision. It puts boundaries or limits to it. The statement of problem states objectives, the methods of the study and the population on whom the result is to be generalized. It directs all research procedures.

11.2 Defining purposes/ objectives

There is need of a justification for purpose of the study. The purpose of the study should base on rationale and presented clearly. It will help to get approval.

11.3 Defining terms

The terms used in the study need to be defined clearly, so that their meaning will clearly understandable to the researcher and readers of the report. To prevent any confusion, all the terms used in the study should be operationally defined.

11.4 Review of literature

Research should base on previous knowledge/ evidence/ practice etc. As soon as we identify the problem, we need to review the problem related literature to find out what has been already done and what need to be done on the identified problem.

11.5 Identifying assumptions

Assumptions are beliefs based on to be true but not necessary. Each scientific research is based on assumptions, so these should be stated clearly. Before starting the research, the researcher searches certain things in other studies; they assume or believe to be true. These beliefs are called assumptions. These assumptions are not tested in this study but researcher feels that there are sufficient evidences in support. Assumptions need to be selected after a thorough review of literature. These are perceptions based on findings of the other study, systematic observation and general theories.

The researcher should use a sound sense of judgment to choose the research assumptions. The researcher should identify the limitations or weakness of the study. These limitations are uncontrolled variables (extraneous) that may affect the result of the study.

11.6 Operational definition- It allows the researcher to describe in a specific way, they mean when they use a certain term. The researcher clarifies and defines terms that how he/she will be observed and measured in the study.

11.7 Stating hypotheses

The word “hypothesis” is used in singular as while “hypotheses” are used in plural form. Hypothesis is considered a statement of the expected outcome or result. It is also called an educated guess by researcher, proposes to test or enquiry. It provides a statement about a specific relationship which is not tested yet. It is based on logical-rationale and has empirical possibilities for its testing.

Normally, we observe four elements in a hypothesis:

- (i) Dependent and independent variables.
- (ii) Relationship between dependent and independent variables.
- (iii) Direction of change, i.e. stating increased or decreased, more or less, higher or lower.
- (iv) It describes the population or subject being studied.

Treece and Treece described following characteristics of a well stated hypothesis

- (i) It is testable.
- (ii) It is logical.
- (iii) It is directly related to research problem.
- (iv) It is factually and theoretically based.
- (v) It states relationship between variables.
- (vi) It is stated in such a form that can be accepted or rejected.

11.8 Selecting research design

The research design includes selecting research approach, planning for data processing, planning for data analysis and planning for interpretation.

11.9 Planning for data collection

(i) Identify the population

Population is a set or a group of individuals or objects that possess some traits or characteristics in which, the researcher has some interest. A target population is a set or a group of individuals or objects that researcher wants to generalize the study findings. Accessible population is a set or a group of individuals or objects that actually available for the study by the researcher.

(ii) Selecting the sample and sampling technique

A sample is a part or an amount of the population to represent the entire population. It may be group of items or objects or people. There are several probability and non-probability methods of sample selection which will be described in further chapters.

(iii) Setting of the study- It is physical location and the conditions in which data collection takes place in the study. The selection of an appropriate set up is important because the set up can influence the way people behave or feel and how they respond. The researcher needs to decide where the interventions will be implemented and from where the data will be collected.

(iv) Identifying/ Developing/ research tool- The researcher decides or selects the data collection technique. After it, he/she is to find out/select the data collection tool suitable to the research study. There may be two options- he/she may

finds out /selects the data collection tool suitable to the research study, then he/she may proceed for next steps, if he/she does not find out /select the data collection tool; he/she has to construct/ design or develop data collection tool suitable to the research study.

(v) Ethical consideration -The word "ethics" refers to rules of conduct and a system of principles. It covers whole field of moral science, e.g. social science, political science, law, jurisprudence etc.

Issues of ethical behaviour are important for health professionals. A passage of the oath of Hippocrates indicates that the information seen or heard in the course of professional behaviour should be held secret. The International Council of Nursing (ICN) says that "Inherent in nursing is respect for life." Both statements are referring to privacy and confidence of client.

As a nursing researcher, we collect data from the clients which he/she may consider as a type of intrusion in personal life. We may find sometimes that the client would say "some information I do not want to share."

(a) Rights of human subjects- As a nursing researcher, we need to respect the rights of our subject that participating in the study. A researcher will not compel the subject. The subject must aware about the research. This term is called as informed consent. After seeing all the factors, they may agree to participate or not in the study.

The researcher needs to guarantee of confidentiality and anonymity. He/she should also assure that the participant is free to respond or to take participation voluntarily. Taking photographs, use of video and tape recorder or mobile recording are the activities which may be included in privacy. If there are some clear-cut guidelines on research ethics, it should be followed clearly but if there are no clear-cut guidelines about it, then the issue of privacy and anonymity is more important. Sometimes, it is emphasized that respondents can demands to know the result of the study, as the respondents are also consumers. In such cases, findings of the study may be informed to them.

(b) Freedom from harms- The researcher may include subjects in painful experience, drugs or treatment etc. In these types of cases, researcher should take step to protect subject from physical or mental harm. It is seen that a good explanation and consent prior to participate in research, prevents unpleasant experience. When the researcher observes any individual subject under painful experience or unusual emotion, he must be taken to calm the subject and data collection procedure should be stopped.

(vi) Planning for data analysis- Data analysis includes categorizing, ordering, manipulation and breaking down of data, which helps to test the hypotheses and to obtain the answer of research question. It is based on the research objec-

tives. There may be two types of analysis-descriptive analysis/static and inferential analysis/static.

(vii) Conducting a pilot study- A pilot study is done on a small sample on trial version of the full study before the actual data are collected. It is done on the people who have similar characteristics to the population of actual study. It is to find out the feasibility of the study. It is used to test the tool and select the appropriate methodology. It also identifies the potential problems in research study.

11.10 Data collection, analyzing and interpretation of results

After satisfaction of pilot study, the data of actual and full research data are collected. The data are the part of information that is collected during study with the help of research tool. The data are collected, organized, analyzed and interpreted.

11.11 Writing the report, critiquing, publication and application of results

When the data are collected, organized, analyzed and interpreted, the findings of the research study are written as report. It is criticized and prepared for publication accordingly i.e. for thesis or dissertation, article for journals or newspapers or mass media etc. If anyone finds research useful or applicable, then it is applied accordingly.

Let us check our progress-5

Discuss ethical consideration in nursing research.

12. Let us sum up

Research is a systemic process of investigation to invent new knowledge. Research problem is a mysterious condition about an area of concern, Research question is a statement of specific query to which, researcher wants to search an answer of the question. The clearly defined/written objectives are very important to get an answer or a solution of a problem. There are two types of research objectives- general and specific. The specific purposes of research are -description, exploration and prediction. The scientific method is an objective, logical and systemic that aim to build general knowledge about natural phenomena. The word "ethics" refers to rules of conduct and a system of principles. It covers whole field of moral science, e.g. social science, political science, law, jurisprudence etc. Issues of ethical behaviour are important for health professionals. As a nursing researcher we collect data from the clients whom he/she may consider as a type of intrusion/ invasion/ encroachment in personal life. Rights of human subjects and freedom from harm are two important ethical issues. The research process is described in three stages- planning, implementing and communicating and applying results. These stages are further divided in more steps. In the next chapter, we will talk about review of literature.

13. Answers to let us check our progress

Let us check our progress-1

(i) *Write the meaning of research question.*

Please refer para 4.2, page 4.

(ii) *State the criteria of a good research problem.*

Please refer para 4.4, page 5.

Let us check our progress-2

(i) *Define research.*

Please refer para 5.1, page 5.

(ii) *Define nursing research.*

Please refer para 5.2, page 6.

Let us check our progress-3

(i) *Describe characteristics of research objectives.*

Please refer para 6.3, page 8.

(ii) *Describe purposes research.*

Please refer para 7, page 8.

Let us check our progress-4

Enumerate the steps of scientific methods.

Please refer para 8.3, page 9.

Let us check our progress-5

Discuss ethical consideration in nursing research.

Please refer para 11.9(V), page 13.

CHAPTER-2

Review of Literature

Structure

1. Introduction
2. Objectives
3. Review of literature
 - 3.1 Meaning
 - 3.2 Purposes
 - 3.3 Types
 - 3.4 Source
 - 3.5 Method/steps of review
4. Let us sum up
5. Answers to let us check our progress

1. Introduction

Review of literature is one of the most important step in the planning stage of any research. Although, it is very laborious work but is essential for success of the research process. Before starting any research study, a review of previous study and experience related to the desired topic or field is to be done under review of literature. It provides a direction toward a futuristic need. It helps to find out the research problem, objectives and methodology for proposed study. It is also applied in field of nursing research. It is so important component of research process that in any research study or dissertation, one chapter is written of review of literature. The researcher needs knowledge, expertise and skills for reviewing the related literature. In this chapter, we will study the meaning, importance, purposes, types, sources and the methods/ stages of review of literature. We will know the important points to be remembered in reviewing the literature.

2. Objectives

After studying this chapter, we will be able to:

1. Understand the meaning of review of literature.
2. Know the importance of review of literature.
3. Explain the purposes of review of literature.
4. Describe the types of review of literature.
5. Discuss the sources of review of literature.
6. Explain the methods/ stages of review of literature.
7. Remember the important points in reviewing the literature.

3. Review of literature

3.1 Meaning

Review of literature tells us about the work already done, who contributed in it, which

of the problems has already solved, what methodology was used in on the particular topic or field of interest. It also shows the gap between the works done. It provides the topic/ area for further research. It is systematic, scholarly and critical review of relevant literature. It includes books, references, periodicals, dissertations, journals, articles, reviews, abstracts, critics, research reports etc.

3.2 Purposes

The literature is reviewed before conducting a research study for many purposes. Some of important purposes are as below:

- (i) To determine what is already known and get background for understanding about the topic of interest.
- (ii) To find out the recommendations in previous study for future research.
- (iii) To take help in planning for conceptual framework, research methodology, design, selection/ framing of tool, sample/sampling and plan for statistical analysis.
- (iv) To find out the related data that could be used for supporting present findings, discussion and conclusions.
- (v) To minimize the unintentional duplication about the problem of study.
- (vi) To increase the possibilities of new and useful research problem for professional practices.

Let us check our progress-1

- (i) *Write the meaning of review of literature.*
- (ii) *Explain the purposes of review of literature.*

3.3 Types

Following are the common types of literature:

- (i) **Subject books-** These can provide information for background, theory and methods of the research.
- (ii) **Grey literature-** It is any information that not produced by commercial publishers, they don't have International Standard Book Number (ISBN) or International Standard Serial Number (ISSN). It includes newspaper articles, unpublished theses, conference papers etc.
- (iii) **Official publications-** It can be used as valuable source for background and contexts. It can be a used as secondary source of data.
- (iv) **Research reports of funding agencies-** Copies of any research is also kept with the funding agencies. Information related topics/ projects may be gain from there.
- (v) **Research reports presented in conferences-** The conferences are conducted at various levels i.e. International, National or University level/local levels. Researchers present their research papers in the conferences. We can get the needed information during presentations.

- (vi) **Writing aids-** This type of literature useful in research process. It also improves the linguistic style of writing work. It includes yearbooks, encyclopedias, bibliographic works, dictionaries etc.
- (vii) **Journal articles-** Articles taken from the journals considered credible, targeted for academic persons. These are available on regular basis, because journals are published in regular periods/ interval. We can get current copies as well as archives of research journals.

Let us check our progress-2

Enlists the types of review of literature.

3.4 Source

The sources of review of literature may be classified as under:

- (A) **Print literature-** These literature sources are available in libraries of universities, colleges, national libraries, research clubs etc.
- (i) **Books-** These are used as primary source in concerned areas of interest. These are direct sources of review of literature.
- (ii) **Journals-** Journals provide original research reports, detailed methodology and results related areas of interest. Journals are available in two forms: non-referred and referred. In non-referred journals, articles are not reviewed by the research experts before publication. But in referred journals, articles are reviewed by the research experts before publication. The experts reject the several manuscripts of articles, if not found at the level of perfectness. The selected manuscripts are published as articles in journals.
- (iii) **Abstracts-** These are the summary of research study in various fields. These provide information as the title, name of authour, journal pagination etc. about research article. These are source of up-to-date information but don't provide detailed of methodology.
- (iv) **Post-graduation/ doctoral dissertations-** These are important source of review of literature related to field. These are found in college or university libraries. Researcher can select according to need. Most of the dissertations have a chapter of review of literature.
- (v) **Reviews-** These provide brief information about previously published articles. These are done for a particular period of time. These are published in journals, handbooks, yearbooks and encyclopedias. The articles of interest are selected and organized by the reviewers. They criticize their findings, add their own suggestions and conclude the materials. These are beneficial for those researchers who want to find material at one place.
- (vi) **Indexes-** These provide only titles of research study. These titles are found categorized and arranged alphabetically. These are considered only supplementary sources.

- (vii) Periodicals
- (viii) Encyclopedias
- (ix) Research articles, reports, documents, critics etc.
- (x) Conferences/ seminars/ workshop/ meeting proceedings or reports.
- (xi) Manuals
- (xii) Bulletins
- (xiii) Annual reports
- (xiv) Official records.

Print literature further may be classified as follows:

(a) Primary sources

Primary sources provide direct information conducted/ prepared by original researchers themselves. These are original descriptions and research reports. Mostly these are published materials. These sources are preference in literature of review by a good researcher.

(b) Secondary sources

Secondary sources provide indirect information conducted/ prepared by some one others than original researchers. These are the comments, interpretations and summaries of original work done by others. These sources are used when primary literature of review is not available.

(B) Electronic literature

The existence of computer based programmes revolutionized every sector of life. Today is the age of internet; life becomes difficult when internet supply interrupted. Electronic literature search may be very useful but sometimes, we find it time consuming and predictable. There are many websites, sometimes they may overload the information, it may result confusion. Like other areas, it also affected the review of literature. Electronic literature may be searched through search engine and database sources. A search engine is a collection of software programmes. It connects the information from web and put it in a format. It may be accessed visually on screen at on-line library or in readable written or in downloadable format. A database is an organized body that has related information. This information is arranged for speed of access and retrieval. A database is storage location, where information is stored, cataloged, maintained and uploaded systematically. There are two types of data base available, first bibliographic, that gives only direction on where to find the information and the second is the full-text data base, contains the information in full research article in downloadable format. It provides articles and bibliographies of original researcher. Some professional/ academics/ research societies and/or associations provide support to researchers. Researchers may find many use-

ful information regarding their field of interest.

Following are the useful and relevant nursing databases:

- (i) CINHALL (Cumulated Index to Nursing and Allied Health Literature)
- (ii) MEDLINE (Medical Literature Analysis and Retrieved System Online)
- (iii) PubMed
- (iv) British Nursing Index
- (v) MedlinePlus
- (vi) Nursing and Health Science
- (vii) Registry of Nursing Researcher
- (viii) Cochrane database of system reviews
- (ix) ERIC
- (x) PsychINFO
- (xi) Dissertation abstract online
- (xii) Online journals
- (xiii) Health STAR (Health services technology administration and research)
- (xiv) Radix (Nursing and managed care database)

(C) Research guides/supervisors

These are the real source on subject/ topic/ research problems. Their suggestion and advise should be taken seriously. They take parts in continuous consultation in many areas with many students.

Let us check our progress-3

Discuss the sources of review of literature.

3.5 Method/writing process/ steps of review

Writing of review of literature is an important thing in any research. The following steps may be followed:

- (a) How old the literature be reviewed-** The literature may be reviewed of any age. The research is a cumulative activity. Each generation learns from previous and current research depends on the work done by previous researchers. In most of disciplines some work is so significant that it is valued in all age. The researchers should always take precaution in choosing the old. They should take consideration when they are confident.
- (b) Reading of good and relevant reviews-** Researcher should give attention on the literature reviewed from the several journals and theses. Out of them, he/she should select some good and relevant reviews. He/she may take help from their supervisors.
- (c) Writing critical annotations during reviewing process-** Researcher should start selecting and organizing the annotations of reviews by issue of concern,

theme and limitations. It will help the researcher to develop own thinking or arguments.

- (d) **Developing a needful structure-** Researcher should develop structure according to needs. It is always subject to modification according to new thinking and needs.
- (e) **Preparing of index cards-** When the required literature is reviewed, the researcher should prepare their own review and abstract on the index card. In journals, most of the articles has abstract of about 150 words. After review, the researcher should make abstract in own words. Mostly the abstract is seen in three parts. The first part includes the purpose and hypotheses, the second part consists of methodology and the third part includes the finding and conclusion of the study. The researcher should never believe on his/her memory; therefore, all the important and relevant detail should be noted carefully in the index card.
- (f) **Writing purposefully-** Review of literature may be done without certain objectives. But when any researcher starts to write the reviewed literature in formal style, he should know that why and what he wants to write. It means he should have some objectives or aim of review of literature.
- (g) **Developing self-arguments-** Researcher should develop own ideas to support his arguments on the topic with the help of the reviewed literature.
- (h) **Making review of literature a continuous process-** Researcher should make review of literature a continuous process, it should relate to research problem, theories and methods. These are helpful in setting the parameters of the review of literature. It becomes a cyclical process.
- (i) **Getting feedback continuously-** As soon as he writes the review, he/she should forward it to the supervisors and experts for their feedback. It provides a chance for rethinking and modification of ideas being included in the writing process.
- (j) **Be ready for redrafting-** After getting feedback from supervisors and experts, researcher should redraft the review with reasonably supports his arguments.

Let us check our progress-4

Describe the steps of review of literature.

4. Let us sum up

Review of literature is one of the most important steps. It is very laborious work but is essential for success of the research process. It helps to find out the research problem, objectives and methodology for proposed study. The researcher needs knowledge, expertise and skills for reviewing the related literature. Review of literature tells us about the work already done, who contributed in it, which of the problems has already

solved, what methodology was used in on the particular topic or field of interest. It also shows the gap between the works done. It includes books, references, periodicals, dissertations, journals, articles, reviews, abstracts, critics, research reports etc. There are many purposes for reviewing the literature before conducting a research study. The sources of review of literature may be classified many types. Writing of review of literature is an important thing in any research. In the next chapter, we will learn the research approaches and research designs.

5. Answers to let us check our progress

Let us check our progress-1

(i) Write the meaning of review of literature.

Please refer para 3.1, page 17.

(ii) Explain the purposes of review of literature.

Please refer para 3.2, page 18.

Let us check our progress-2

Enlists the type of review of literature.

Please refer para 3.3, page 18.

Let us check our progress-3

Discuss the sources of review of literature.

Please refer para 3.4, page 19.

Let us check our progress-4

Describe the steps of review of literature.

Please refer para 3.5, page 21.

CHAPTER-3

Research Approaches and Designs

Structure

1. Introduction
2. Objectives
3. Research approaches
 - 3.1 Meaning
4. Types of research
 - 4.1 Historical research
 - 4.2 Survey research
 - 4.3 Experimental research
5. Research design
 - 5.1 Meaning
 - 5.2 Types -qualitative and quantitative
6. Let us sum up
7. Answers to check our progress

1. Introduction

Research approaches and research design are two terms but these are used interchangeably. Research design is a broader term. Research approaches includes in the design. In this chapter, we will study the meaning, types of research approaches and research designs.

2. Objectives

After studying this chapter, we will be able to:

1. Understand the meaning of research approaches and research designs.
2. Describe the types of researches.
3. Discuss the research approaches.
4. Explain the research designs.

3. Research approaches

3.1 Meaning

Research approach is a plan and procedure. It includes the steps and assumptions. It has method of data collection, analysis and interpretation. It is based on the nature of the research problem. Based on the nature of the research problem, it may be classified in various types. It is overall theoretical/ philosophical strengthening of the study. Research design is doing part of the research process.

4. Types of research

Based on the timing, there are three types of researches i.e. historical (past), survey

(present) and experimental (future).

4.1. Historical research

Historical research is related with identification, location, evaluation and synthesis of data from the past. It discovers the past and relate with the present and future. According to Leininger (1985) “Without a past there is no meaning to the present, nor can be develop a sense of ourselves as individuals and as members of groups.” This type of research is more difficult than other type of research. The data of historical research usually found in documents (printed material, oral reports found in libraries, archives or personal collection) or in relics.

(A) Sources

Primary sources- These are firsthand information or direct information, e.g. oral history, written records, diaries, eye witness, pictorial sources, physical evidence.

Secondary sources- These are the second or third hand information or indirect information of event or experience, e.g. text books, encyclopedia, other reference books or interpreting the contents.

(B) Data synthesis and analysis: Data collected from the above mention sources, should be organized. Relative value of sources used, must be weighed. There should make an objective judgment regarding sources. Formulation of interpretation is made in the light of hypothesis and question.

(C) Evaluation of data: The evaluation of data may be categorized in two types:

(i) External criticism: It is concerned with authenticity or genuineness of data. The researcher must decide the validity and relevancy to the research problem.

(ii) Internal criticism: It is to examine the accuracy. The researcher must decide the reliability.

(D) Steps of conducting historical research

- a. Formulation and defining the problem.
- b. Formulation of specific objectives and hypotheses
- c. Plan for data collection methods, tools and sources
- d. Collection of data
- e. Evaluation of data
- f. Interpreting the findings
- g. Analyzing the results
- h. Communicating the results and present status

4.2. Survey research

Survey research focuses on the present. It is a non-experimental method. It is used to find out characteristics, opinions, knowledge, attitude etc. of the population. In this research, data may be collected in number of ways. Data may be collected by interview and by observations of the incidents according to the objectives.

Questioning is the most common method. It can be used as an interview schedule or self-administered questionnaire. The tool should be valid and reliable.

In survey research, generally data are collected from the sample not from person to person. An appropriate selection of sample size and sampling technique should be used. The probability sampling technique is considered better method for selection of sample.

Purposes- There are three main purposes of survey research- description, explanation and exploration.

- (i) **Description-** The researcher discovers the distribution of certain traits or characteristics in existence.
- (ii) **Explanation-** The researcher examines two or more variables to find out relationship between phenomena.
- (iii) **Exploration-** The researcher explores the situation where there is very little information exists.

Types- There are various types of survey research e.g. descriptive, comparative, evaluative, correlation, longitudinal, cross-sectional, cross-cultural etc.

Advantages

- (i) It is less costly than other type of research, e.g. experimental research.
- (ii) It is a good source of generating hypothesis.
- (iii) It provides insight into a situation or new insight is developed.
- (iv) It has higher degree of representatives, if proper sampling technique is followed.
- (v) It helps to identify the variables.
- (vi) It may be used to get information about to understand the individual's/ group's behavior, e.g. values, expectations, relationship etc.
- (vii) Its research findings may be used in replication of the study.

Limitations

- (i) The extraneous variables are not under control.
- (ii) People often not express their true reaction to the question related to research.
- (iii) Unreliable and invalid tools may affect the results.
- (iv) Verbal answers are not reliable.

4.3. Experimental research

Experimental research is related with cause and effect relationship. Any experimental research need following three properties:

- (i) **Manipulation-** The researcher manipulates some variables.
- (ii) **Control-** The researcher uses controls over experimental situation.
- (iii) **Randomization-** Unreliable and invalid tools may affects the results.

Purposes- Its purpose is to show the cause and effect relationship between variables.

Types- There are four types of experimental research

(A) Pre test-post test control group design- In this design/type:

- (i) The subjects are randomly assigned to group.

- (ii) Pre test is given to both the groups.
- (iii) The experimental group receives the experimental treatment/ intervention and the comparison group receives the routine treatment/ intervention.
- (iv) Post test is given to both the groups.

Pre test-post test control group design is the simplest true experimental design.

Groups	Pre test	Treatment	Post test
Experimental	O ₁	(×)	O ₂
Control	O ₁	-	O ₂

In this design, the researcher selects the population and assign them randomly to either experimental or control groups.

(B) Post test only controls group design- In this design/type:

- (i) The subjects are randomly assigned to group.
- (ii) The experimental group receives the experimental treatment/ intervention and the comparison group receives the routine treatment/ intervention or no treatment.
- (iii) Post test is given to both the groups.

Groups	Pre test	Treatment	Post test
Experimental-1	O ₁	(× ₁)	O ₂
Control-2	O ₁	(× ₂)	O ₂

(C) Solomen four group design- In this design/type:

- (i) The subjects are randomly assigned to one of four groups.
- (ii) Pre test is given to two of the groups, experimental and control.
- (iii) Two of the groups which are experimental, receives the experimental treatment/ intervention, whereas two of the groups, as control receives the routine treatment/ intervention or no treatment.
- (iv) Post test is given to both the groups.

Groups	Preintervention	Intervention	Post-Intervention test
1.	Test	Treatment	Test
2.	Test	Control	Test
3.	No-Test	Treatment	Test
4.	No-Test	Control	Test

(D) Pre-experimental design- In this type of design, the researcher has little con-

trol over the research. There are two types of designs; one is for case study and one group pre test-post test design.

Advantages

- (i) It is a powerful method of testing hypothesis of the cause and effect relationship between variable.
- (ii) It has a special controlling property on experiment, it offers greater control than any other research approach that the independent variable has an effect on the dependent variable.
- (iii) The great strength of experiments then lies in the confidence with which causal relationship can be inferred.

Limitations

- (i) Many variables are not amenable to experimental manipulation.
- (ii) A large number of human characteristics cannot be randomly selected.
- (iii) Many situations as in hospitals may not be feasible.
- (iv) Placebo effect is a problem that effect on the dependent variables caused by subject's awareness, that they are special participants under study.

5. Research design

5.1 Meaning: A research design is a framework that has been created to find answers to research questions. It is overall strategy utilized to carry out research study. It defines the study type, data collection methods and a statistical analysis plan.

5.2 Types - There is no uniform classification of research design, each author classify it, in his own style. Generally, research designs are classified into following broad categories

1. Qualitative research design

Qualitative research aims to study things in their natural settings to make sense of a phenomenon in terms of meanings people bring to them. It relies on data obtained by the researcher from first hand observation, interviews, questionnaires, documents and recording made in natural settings. These are mostly non-numerical. It involves collecting and analyzing non-numerical data to understand concepts, opinions etc.

- (i) **Phenomenological research design:** It examines human experiences. It is done through the descriptions provided by people involved.
- (ii) **Ethnographic research:** It collects the information from certain cultural groups. It is obtained by living with people of those groups. It may get from their informants, who have most knowledge about the selected culture.
- (iii) **Grounded theory design:** In this design, the analysis and development of theories are done after actual data collection. These data are used for explanatory purposes. It is to explain a phenomenon that is not well understood. It was introduced to legitimate qualitative research.
- (iv) **Case studies:** It is a detailed and in-depth study of a specific subject. It is

commonly used in educational, clinical, business and social research, e.g. a person, place, group, event or phenomenon.

- (v) **Historical research design:** It studies past events and attempt to interpret the facts and explain the cause of events and their effect in past. It might help in explaining current events and expecting future events.
- (vi) **Action research:** It is a design in which the researcher and a client collaborate. This collaboration is done for diagnosis of the problem and development of solutions. It is used to improve practices and study the effect of the action that was taken.

2. Quantitative research design

Quantitative research is a process of collecting and analyzing numeral data. It is generally used to find averages, patterns, predictions and cause-effect relationships between the study variables. Its purpose is to attain knowledge and understanding of social world. It is used to observe situations or events that effect people. It produces objective data. These data can be communicated through statistics and numbers. Following are the types of quantitative research designs:

I. Experimental research design

(A) True experimental research design: Following are main characteristics/ attributes of this design:

- (i) Manipulation of independent variable.
- (ii) It is done in presence of control group.
- (iii) Randomization is done.

Following are the types of the true experimental research design:

- a. Post test only control design.
- b. Pre test-Post test control group design.
- c. Solomon four-group design.
- d. Randomized block design.
- e. Crossover design.
- f. Latin square design.

(B) Quasi-experimental research design: Following are main characteristics/ attributes of this design:

- (i) Manipulation of independent variables.
- (ii) It is done in absence/presence of either randomization or control group.

Following are the types of the quasi-experimental research design:

- a. Non-randomized control group design.
- b. Time-series design.

(C) Pre-experimental research design: Following are main characteristics/ attributes of this design:

- (i) Manipulation of independent variables.
- (ii) It has limited control over extraneous variables.

(iii) No randomization and no control group.

Following are the types of the quasi-experimental research design:

- a. One-shot case design.
- b. One group pre test-post test design.

II. Non-experimental research designs

(A) Descriptive research design: It is an accurate description of characteristics of an individual, situation or group. It has no control or manipulation. It occurs in a natural setting.

Following is the type of the descriptive non-experimental research design:

- a. Uni-variant descriptive research design: These studies undertaken to describe the frequency of occurrence of a phenomenon rather than to study relationship
 - b. Exploratory descriptive research design: It investigates the phenomenon and its related factors about which very little is known.
 - c. Comparative descriptive research design: It compares the phenomenon in two or more groups.
- (B) Correlational/ ex post facto research design:** It studies the relationship between two or more variables in a natural setting without manipulation or control (cause and effect relationship)
- a. Prospective research design: It studies the relationship from cause to effect.
 - b. Retrospective research design: It investigates the relationship from effect to cause.
- (C) Developmental research designs:** It studies the phenomenon with respect to the time.
- a. Cross-sectional research design: It investigates the phenomenon only at one point in time.
 - b. Longitudinal research design: It studies the phenomenon only at more than one point in time.
- (D) Epidemiological research designs:** It investigates the distribution and causes of diseases in a population. It is known as epidemiology.
- a. Case-control research studies/design
 - b. Cohort research studies/design
- (E) Survey research design:** It investigates the self-reported data collected from sample with the purpose of describing population on some variable of interest.

III. Other research designs

- (A) Methodological research design:** It investigates to develop, test and evaluate the research instruments and methods.
- (B) Meta-analysis research design:** Quantitatively combining and integrating the findings of the multiple research studies on a particular topic.
- (C) Secondary data-analysis research design:** In this type of research, the data

collected in one research study is analyzed by another researcher. It is done to test new hypotheses.

- (D) **Outcome research design:** It involves the evaluation of care practices and systems in place.
- (E) **Evaluation research design:** It involves the judgment about success of programmes, practices, procedures or policies.
- (F) **Operational research design:** It involves the study of complex human organizations and services. It is to develop new knowledge about institutions, programmes, use of facilities and personnel. It is to improve working efficiency of an organization.

IV. Mixed methods

This type of research design is prepared by mixing the above both designs according to need.

Except the above classification of research design, there may be other classifications also. The following are examples of research designs:

(i) **Based on the number of contacts with the study population:** It can be classified in following three groups:

A. Cross-sectional studies design: It is also known as one-shot design studies. It aimed at finding out the prevalence a problem, attitude or issue. It is done by taking a cross-section of the population. It is a simple design for study. We decide what we want to find out, identify the study population, select a sample and contact our respondents to get the required information. This design can be used to find out the immunization need of a community against corona virus. We can also find out the attitudes of our nursing students towards the facilities available in their clinical area. It is easy to analyse.

Demerit: It has a major demerit that it cannot measure change. The change can be measured at least two data points in same time and on the same population.

B. Pre and Post studies design: It is also known as the before-and-after design studies. It can measure change in a problem, attitude, situation or issue. It is used for measuring the impact or effectiveness of a programme. It can be described a two sets of cross-sectional data collection points on the same population to find out the change in the problem, attitude, situation or issue between two points of time. The change is measured by comparing difference in the phenomenon before and after the intervention. It is carried out by same process as a cross-sectional study but it contains two cross-sectional observations, the second observation is taken after a certain period. It may be either an experimental or a non-experimental. It is used in evaluation studies. This design can be used to find out the effect of a tobacco chewing awareness programme on the knowledge and use of tobacco chewing among college students. We can also find out the impact of maternal and child health services on the infant mortality rate.

Demerit:

- ❖ It is more expensive as well as difficult to implement, because we have to collect two set of population. If we are using experimental design, we have to wait until our intervention is completed.
- ❖ Sometimes, some samples may left out the area, who participated in pre-test.
- ❖ This design measures the total change, we cannot explain the cause, responsibility of change; independent or extraneous variables.
- ❖ Sometimes the research tool learns the sample/ respondents. During pre-test respondents learn many things, which are helpful to them in post-test.

C. Longitudinal studies design: The pre and post study design is capable to measure change in a problem, attitude, situation etc. but it is not capable to measure the pattern of change. Longitudinal studies are used to study the pattern of change in relation to time. These studies are also helpful to collect factual information on a continuous basis. We may use it to find out the trends in changes in incidence of corona virus infection of a population. In these studies, the study population is visited a number of times at a regular intervals over a long time to collect the required information. These intervals are not fixed; it may a week to longer than a year. The data are collected from the same population but respondents may differ.

Demerit:

- ❖ Same as that of pre-test and post-test.
- ❖ If the same respondents are contacted frequently, they start to know what is expected of them and they may answer without thought and may lose interest.

(ii) Based on the reference period: It refers to the time-frame in which a study is exploring a problem, event or situation. It can be classified in following three groups:

A. Retrospective studies design: It investigates a problem, event, issue or situation that has happened in the past. It is conducted either on the basis of data available for that period or on the basis of recall of the event or situation by the respondents. This design can be used to find out the strategy adopted at primary health care level about eradication programme on chicken pox.

B. Prospective studies Design: It refers to prevalence of a problem, event, issue or situation in the future. It attempts to show the outcome of a problem or what likely to happen. This design can be used to find out the effect of parent-teacher-student meet on academic achievement of student in our university.

C. Retrospective-prospective studies design: It focuses on the past trend in a phenomena or problem and studies it into future. The data collection is done retrospectively from the existing records before the intervention, the intervention is given and then data are collected to find out the effect of the intervention. We can find out the impact of maternal and child health services on the infant morbidity and mortality rate.

(iii) Based on the nature of investigation

A. Experimental studies design: Imagine we want to test the effect of discussion

teaching method on the level of understanding of students. When we start a study from cause to establish the effects, it is called as an experimental study.

B. Non-experimental study design: When we start a study from effects to find out the cause, it is called non-experimental study.

C. Semi-experimental study design: It has properties of both experimental and non-experimental study design. One part of the study may be non-experimental and the other part of the study may be experimental.

(iv) Other designs are also in existence, e.g. action research, cohort studies, trend studies, panel studies, feminist research, the cross-over comparative experimental design, blind studies, double-blind studies, case studies etc.

Let us check our progress-1

Write the meaning of research approach.

Let us check our progress-2

Describe the types of research.

Let us check our progress-3

(i) Write the meaning of research design.

(ii) Explain research designs.

5. Let us sum up

Research approaches is a plan and procedure that includes the steps and assumptions to detailed method of data collection, analysis and interpretation. It is overall theoretical/ philosophical strengthening of the study. Research design is doing part of the research process. Based on the timing, there are three types of researches i.e. historical (past), survey (present) and experimental (future). There is no uniform classification of research design, each authour classify it, in his own style. Generally, research designs are classified into qualitative research design, quantitative research design and mixed design. In the next chapter, we will like to discus about sampling and data collection.

6. Answers to let us check our progress

Let us check our progress-1

Write the meaning of research approach.

Please refer para 3.1, page 24.

Let us check our progress-2

Describe the types of research.

Please refer para 4, page 24.

Let us check our progress-3

(i) Write the meaning of research design.

Please refer para 5.1, page 28.

(ii) Explain research designs.

Please refer para 5, page 28.

CHAPTER-4

Sampling and Data Collection

Structure

1. Introduction
2. Objectives
3. Sampling
 - 3.1 Concepts
 - 3.2 Purposes or need
 - 3.3 Advantages
 - 3.4 Disadvantages
 - 3.5 Characteristics of a good sample
 - 3.6 Suitability
 - 3.7 Sampling process
 - 3.8 Types of sampling techniques/ methods
4. Data collection
 - 4.1 Concept
 - 4.2 Techniques
 - 4.3 Tools
 - 4.4 Pilot study
 - 4.5 Data collection procedures
5. Data organization
6. Data presentation
7. Let us sum up
8. Answers to let us check our progress

1. Introduction

Sampling is a base of the research process. Researcher selects a sample from the population because it is not possible to include whole population in the research study. Researcher uses sampling techniques and research tools in data collection. After collection of data, these are organized and presented in a meaningful form. In this chapter, we will know the concepts about population, sample and sampling. We will learn the advantage, disadvantages of sampling, the characteristics of a good sample and types of sampling techniques. We will discuss about collection, organization and presentation of data.

2. Objectives

After studying this chapter, we will be able to:

1. Understand basic concepts used in sampling.
2. Know the advantage and disadvantages of sampling.
3. Describe the characteristics of a good sample.
4. Explain the types of sampling techniques/ methods.

5. Discuss the needs/ purposes of sampling.
6. Understand basic concepts used in data collection.
7. Know the classification of data collection techniques.
8. Explain the data collection techniques.
9. Describe the qualities of the research instrument or tool.
10. Discuss pilot study.
11. Know the procedures related to data collection, data organization and data presentation.

3. Sampling

3.1 Concepts

Universe/population- In a research study, the participants are collectively called population. The well-specified and identifiable group is known as a population or universe. A population may be finite (where all the members can be easily counted) or infinite (where the size is unlimited and its members cannot be counted). Population is a set or a group of individuals/objects that possess similar traits or characteristics of researcher's interest. It is not restricted to human beings only. A target population is a set or a group of individuals/objects that researcher wants to generalize the study findings. Accessible population is a set or a group of individuals/objects that actually available for the study by the researcher.

Sample- The well-specified and identifiable group is known as a population or universe and the selected number of persons or objects is called a sample. A sample is a part or an amount of the population to represent the entire population. It may be group of items or objects or people. It is a smaller, manageable version of a large group. It is a specific group that we will collect data from.

Sampling- The process of selecting sample or a part or an amount of the population to represent the entire population.

Sampling frame- Sampling units are basis of sampling process. A list of all sampling units is called sampling frame. It consists of a list of items from which sampling is to be drawn. When it may not possible to draw a sample directly from a population, then, the researcher constructs the sample frame according to the purpose or need of the study.

Sampling design- It is a plan for obtaining a sample from the sample frame. It is a procedure to select some sample units. A good sampling design must possess the following characteristics:

- (i) It must provide a sample which would be the true representative of the population.
- (ii) It should be such that it may result in small sampling error.
- (iii) It must allow for controlling systematic bias in the sampling.
- (iii) It must be usable in light of the funds available for the research study.
- (iv) It must be such that the results of the sample study can be applied in general, for the population with a reasonable confidence level.

There are two types of sampling designs- probability sampling and non-probability sampling.

Sampling error- A sampling error is the difference between a parameter and estimate of the parameter which derived from a sample. It is the difference between a parameter and statistic.

Sampling distribution- It is a theoretical or idealized distribution, which is obtained if all possible samples of the given size are taken from the population.

Confidence level- It indicates the likelihood that the answer will fall within the precision range. It is the expected percentage of times stipulating that the actual value will fall within the stated precision limits, e.g. if we take a confidence interval of 95%, there are 95 chances out of 100 that the sample results present the true characteristics of the population within a specified precision range.

Significance level- It indicates the likelihood that the answer will fall outside the precision range. It is the expected percentage of times stipulating that the actual value will fall outside the stated precision limits, e.g. if we take a confidence interval of 95%, then, there are 95 chances out of 100 that the sample results present the true characteristics of the population within a specified precision range against 5 chances out of 100 that it does not. Here the significance level is $100-95=5\%$ or 0.05.

Probability- It means "how likely something is to happen." It is used as equivalent to the relative frequency. It may be expressed in terms of a fraction or in decimal numbers.

Parameter- It is a measure or a numerical value based on the entire population.

Statistic- It is a measure or a numerical value based on a sample.

Let us check our progress-I

Write the meaning of followings:

- (a) Population
- (b) Sample
- (c) Sampling
- (d) Sampling design
- (e) Sampling framework
- (f) Confidence level
- (g) Significance level
- (h) Parameter

3.2 Purposes or needs of sampling

Followings are the purposes or needs of sampling:

- (a) When population contains many members, it is the only way to get information.
- (b) It saves the time, money, material, efforts and resources etc.
- (c) It is conducted by trained and experience researchers, so it provides accuracy in measurement.
- (d) It enables the researcher to estimate the sampling errors. It helps in to find out

information about some characteristics of the population.

- (e) It helps in making generalization after research study because it makes correct and scientific judgment.

3.3 Advantages of sampling

- (a) It is economical in terms of time, money, material, efforts and resources etc.
- (b) It helps in accurate collection of data.
- (c) It facilitates in establishing the close rapport between the researcher and the sample/participants, because sample size is smaller than the population.

3.4 Disadvantages of sampling

- (a) There is a risk of biased sample selection. It occurs due to over or under representation of the population.
- (b) The selection of true representative sample is a difficult task.
- (c) The researcher without experience or skill in sampling may not obtain the accurate result of the research study.

3.5 Characteristics of a good sample- Followings are the characteristics of a good sample:

- (a) A sample should be representative of the population, from where it is drawn.
- (b) A sample should have the same characteristics, attributes of the population, from where it is drawn.
- (c) A sample should be free from sampling biases.
- (d) A sample should be free from sampling errors.
- (e) Size of sample should be adequate and appropriate for generalization after research study.

3.6 Suitability

Sampling methods are suitable in following circumstances:

- (a) Where the scope of investigation is indefinite and unlimited: e.g. determining average weight of newly born baby.
- (b) Where the units under investigation have not much diversity.
- (c) Where the units are likely to be destroyed or consumed during the process of investigation, e.g., testing of ghee, milk or any eatable items, etc.
- (d) Where the researcher has full knowledge of the rules of sampling techniques or takes help of statistician.
- (e) Where the budget is limited or economy of money, time and other human resources is desired.

3.7 Sampling Process

The following steps may include in this procedure:

1. Determining relevant population and parameters

It is initial step of sampling process. We determine of appropriate population and parameters being used to identify sampling units, e.g., we have to conduct a survey on problems of nursing students in Rajasthan session 2020-21. There is need of specification of following things:

Elements: Nursing students

Sampling units: Nursing colleges and nursing students

Extent: Rajasthan state.

Terms: 01.09.2020 to 31.10.2021

2. Selecting appropriate sampling frame

A list of all sampling units is called sampling frame. It consists of a list of items from which sampling is to be drawn. When it may not possible to draw a sample directly from a population, then, the researcher constructs the sample frame according to the purpose or need of the study.

Here, sampling frame is a list of all nursing students.

3. Choosing sample method- Probability or non-probability

Which of the method is chosen- Probability or non-probability

Probability sampling enables the researcher to choose representative sample design.

Non-probability sampling does not enable the researcher to choose representative sample design.

It is often preferred to probability sampling because of convenience and economy.

4. Determination of sample size

Most of the statisticians agree that the minimum sample size to get any kind of meaningful result is 100 but if our population is less than 100 then we really need to survey all of them.

A good maximum sample size is usually around 10% of the population, as long as this does not exceed 1000. For example, in a population of 6000, 10% would be 600.

In a population of 300,000, 10% would be 30,000. This exceeds 1000, so in this case the maximum would be 1000.

Even in a population of 300,000, sampling 1000 people will normally give a fairly accurate result.

Sampling more than 1000 people is not beneficial in accuracy but it will consume more time and money.

Imagine that we want to survey students at a college which has 7000 students enrolled. The minimum sample would be 100.

This would give us a rough, but still useful, idea about their opinions. The maximum sample would be 700, which would give us a fairly accurate idea about their opinions.

In following conditions, we should choose a number closer to the minimum, when

- ▶ We have limited time and money.
- ▶ We only need a rough estimate of the results.
- ▶ We don't plan to divide the sample into different groups during the analysis, or
- ▶ We only plan to use a few large subgroups (e.g. males / females).
- ▶ We think most people will give similar answers.
- ▶ The decisions that will be made based on the results do not have significant consequences.

In following conditions, we should choose a number closer to the maximum, when

- ▶ We have the time and money to do it.
- ▶ It is very important to get accurate results.
- ▶ We plan to divide the sample into many different groups during the analysis (e.g. different age groups, socio-economic levels, etc).
- ▶ We think people are likely to give very different answers.
- ▶ The decisions that will be made based on the results of the survey are important, expensive or have serious consequences.

5. Selecting sample method to be used- Probability or non-probability

Probability sampling includes the followings methods:

- (A) Simple random sampling
- (B) Systematic sampling
- (C) Stratified random sampling
 - (a) Proportionate stratified random sampling
 - (b) Disproportionate stratified random sampling
- (D) Area or cluster sampling
- (E) Multi-stage sampling
- (F) Multi-phase sampling

Non-probability sampling includes the followings methods:

- (A) Quota sampling
- (B) Accidental sampling
- (C) Judgemental or purposive sampling
- (D) Snowball or chain sampling
- (E) Saturation sampling
- (F) Dense sampling

6. Specify the sampling plan

The researcher should indicate that how he will implement the plan.

7. Selecting the sample

The next and final step in the process is to select the sample and sample units. The researcher can start gathering information with the help of research tool.

3.8 Types of sampling techniques/ methods

There are two types sampling techniques/ methods:

(I) Probability sampling techniques/ methods:

Probability sampling techniques/ methods clearly specify the probability or likelihood of inclusion of each element or individual in the sample.

In these techniques/ methods, the size of the parent population from which the sample is to be drawn, must be known to the investigator.

Each individual or element in the population must have an equal chance of being included in a subsequent sample.

The desired sample size must be clearly specified.

The obtained sample from this method is considered representative, so the results found from such samples are worth generalization.

These are also comparable to similar populations to which they belong.

The negative point of this method is that the researcher has only a limited element of the entire population; it may lead to sampling errors.

The small sample shows the greater the sampling error.

The selection of the units for sample is carried out by chance methods such as flipping coins, drawing numbered balls from an urn or through tables of random numbers.

Probability sampling includes the followings methods:

(A) Simple random sampling

It may be defined in which each and every individual of the population has an equal chance of being included in the sample.

In it, also the selection of one individual is in no way dependent upon the selection of another individual,

A random sample may be selected by using any one of the following method:

(i) Lottery method: e.g. if we want to select a sample of ten students from the fourth year of B.Sc. nursing consisting of sixty students.

We can write the name or roll number of each of the sixty students on separate slips (all equal in size and colour) of paper and fold them in a similar way.

These slips may be placed in a box and reshuffled thoroughly.

A blindfolded person may be asked to pick up one slip.

In this situation, the probability of each slip being selected is $1/60$.

Imagine that after selecting the slip and noting the name written on the slip, it again returns to the box.

In this case, the probability of the second slip being selected again $1/60$, but if it does not return the first slip to the box, the probability of the second slip being selected becomes $1/59$.

When an element of population is returned to the population after being selected is called sampling with replacement and when an element of population is not returned to the population after being selected is called sampling without replacement.

We can say that random sampling is one in which all possible combinations of samples of fixed size have an equal probability of being selected.

(ii) Use of random number tables:

If the population size is large and it is not possible through lottery method.

In such situations random number tables are more appropriate.

A random table is a collection of random numbers.

The random numbers are generated through a probabilistic mechanism. Examples of random tables are: Kendal and smith tables, Fisher and Yates tables, Tippett random tables etc.

Advantages of simple random sampling

- (a) It is regarded as the representative of the population from which it was drawn.
- (b) It reflects theoretically all important characteristics and segments of the population.
- (c) It is the easiest and simplest technique of all probability sampling plans.
- (d) It serves as base upon which all other types of random sampling, because this method can be applied in conjunction with all other probability sampling plans.
- (e) Sampling error associated with any given sample drawn can easily be assessed.
- (f) The researcher does not commit classification errors because he need not know thoroughly the population characteristics prior to selection of the sample.
- (g) Accurate mathematical tests can be applied to judge the randomness of the samples.

Disadvantages of simple random sampling

- (a) In many situations, either it is not possible or it is very difficult to make a list of all the items of the universe.
- (b) If the universe is spread to a wider geographical area it is not possible to give equal opportunity to all the units of universe.
- (c) If the universe consists of many heterogeneous groups, simple random sampling is unsuitable.
- (d) It does not ensure that small number elements in the population will be included in the given sample.
- (e) It does not fully exploit the knowledge that the researcher has concerning the segments of the population.
- (f) It has greater sampling error as compared to stratified random sampling.

(B) Systematic sampling

In this type of sampling, a sample fraction is calculated, e.g., we have to select a sample 20 out of 120 units. The sampling fraction (k) is N/n . (where N is the total number of units of population and n is the size of the sample, $k = 120/20 = 6$). A number between 1 to 6 is chosen randomly. Imagine the number thus selected happens to be 4. Then the sample will consist units associated with numbers: 4, 10, 16, 22, 28, 34, 40, 46, 52, 58, 64, 70, 76, 82, 88, 94, 100, 106, 112 and 118.

We are clear that selection of sample by this method is extremely convenient.

Advantages of systematic sampling

- (a) It lies in its simplicity, operational convenience and even spread sample over the population.
- (b) It is relatively a quick method of getting sample of elements/individuals.
- (c) It is statistically more efficient than simple random sampling.
- (d) It makes easy to check if every n^{th} person/number has been selected. By mistake, if the researcher selects 6th number instead of 5th number, sample will not be seriously affected.
- (e) It is very easy to use. It does not need to employ a table of random numbers or

fix quota from each stratum of the population.

Disadvantages of systematic sampling

- (a) It ignores all individuals/elements between every n^{th} person/element/number chosen.
- (b) Sampling error increases if the list is arranged in a particular order.

(C) Stratified random sampling

In this type of sampling the population is divided into two or more strata. The division may be based upon a single criterion i.e. sex, have two strata-male and female. It may be combination of two or more criteria i.e. sex and stream (science students and non-science students), have four strata male science students, female science students, male non-science students and female non-science students. The divided populations are known as sub-populations. These are not overlapping and constitute the whole population. The divided populations in two or more strata, is considered homogeneous internally. A simple random sampling is done for the desired number from each population strata. Stratified random sampling is done for many reasons.

Stratified random sampling is of two types-

(i) Proportionate stratified random sampling-

In this type of sampling, the size of each stratum is proportionate to the population size of the strata when examined across the entire population or we can say that probability of unit being selected from the stratum is proportional to relative size of that stratum in population.

Advantages of proportionate stratified random sampling

- (a) It increases the representativeness of the sample drawn. It ensures that those elements that exist in a few numbers are also included proportionately in the sample.
- (b) Sampling error is minimized because the sample drawn possesses all the necessary characteristics of the parent population.
- (c) It eliminates the necessity of weighing the elements according to their original distribution in the population.

Disadvantages of proportionate stratified random sampling

- (a) It is a difficult technique. It based on an unrealistic assumption. It is assumed that the researcher knows the composition and distribution of population well before the actual sampling starts.
- (b) It is a time-consuming technique because the samples are drawn according to the proportion of each stratum in population.
- (c) It has the probability of classification error. It will affect the validity and reliability of the obtained results.

(ii) Disproportionate stratified random sampling

In this type of sampling, the size of each stratum is not proportionate to the population size of the strata when examined across the entire population or we can say that probability of unit being selected from the stratum is not proportional to relative size

of that stratum in population.

Some of the strata of the population may be over-represented or under-represented.

Advantages of disproportionate stratified random sampling

- (a) It is comparatively less time-consuming than proportionate stratified random sampling because the researcher is not worried about making proportionate representation of each stratum of the population.
- (b) It does not increase the representativeness of the sample drawn. It does not ensure that those elements that exist in a few numbers are also included proportionately in the sample.
- (c) The researcher is able to give weight to the particular groups of elements that are not represented as frequently in the population as compared with other elements.
- (d) It does not eliminate the necessity of weighing the elements according to their original distribution in the population.

Disadvantages of disproportionate stratified random sampling

- (a) It is assumed that the researcher knows the composition and distribution of population well before the actual sampling starts. It cannot be used in a condition where the researcher has no idea about the composition of the original population.
- (b) In this technique, certain stratum of population is over-presented and some other are under-presented in the sample drawn. This situation may introduce some bias in the sample.

(D) Area or cluster sampling

It is a procedure of selection in which the element of the sample are chosen from the population in groups or clusters rather than singly. The clusters are often pre-existing natural or administrative grouping of the population, e.g. geographical divisions of state, district, cities and community etc. are done on a map. Certain number of them is selected randomly, it is called sample. The researcher interviews all elements of randomly selected clusters or groups for data collection.

Advantages of area or cluster random sampling

- (a) It is useful in research of larger geographical areas.
- (b) The researcher need not have the list of the individuals inhabiting a given area.
- (c) Respondents can readily be substituted for other respondents within the same random section.
- (d) It saves time, because the researcher can concentrate his attention in one specific region.
- (e) It saves cost, because the researcher need not travel great distances for interviewing specific individuals residing at random points in a certain geographical region.
- (f) It possesses the trait of flexibility.

Disadvantages of area or cluster random sampling

- (a) It has a high degree of sampling error. It makes sampling much less fruitful.

- (b) It does not ensure that each sampling unit included in an area, the sample will be of equal size.
- (c) There is a little control of the researcher over the size of each cluster. It introduces bias into the sample.
- (d) It is difficult to ensure that the individuals included in one cluster are independent of other randomly drawn clusters. It may be possible that he is available in one cluster today, may travel in another cluster in next day

(E) Multi-stage sampling

Multi-stage sampling is a process of selecting a sample in two or more successive stages, e.g. if we want to know the quality of health services provided by the government of Rajasthan. First of all we select commissioner headquarters randomly, then we select districts randomly, then we select subdivision randomly and finally we select primary health centres or subcentres randomly for collecting related information.

Advantages of Multi-stage sampling

- (a) It leads more precision in the study.
- (b) It is more economic and convenient in comparison to simple random sampling.
- (c) It contributes to better quality in carrying out the final selection.

Disadvantages of Multi-stage sampling

- (a) These are same as that of cluster sampling.

(F) Multi-phase sampling

In Multi-stage sampling, the each successive stage has a different unit of sample where in multi-phase sampling, unit of sample remain the same or unchanged though additional information is obtained from a sub-sample, e.g. if we want to know the quality of health services provided by the government of Rajasthan in a particular city. In First phase, we collect general information from respondents who used hospital services. In second phase, we collect comprehensive information from the same respondents.

Advantages of Multi-phase sampling

- (a) It is economical, less time consuming and requires less efforts.

Disadvantages of Multi-phase sampling

- (a) These are same as that of cluster sampling.

(II) Non-probability sampling techniques/ methods

Non-probability sampling techniques/ methods have not the way of assessing the probability of element or group of elements of the population being included in the sample. These methods provide no basis for estimating how closely the characteristics of a sample approximate the parameters of the population from which the sample had been drawn. It is because samples of these methods don't use techniques of random sampling.

Non-probability sampling includes the followings methods:

(A) Quota sampling

- ❖ It is a non-probability sampling technique.
- ❖ It is apparently similar to stratified random sampling.
- ❖ The researcher recognizes the different strata population.
- ❖ He selects from each stratum the number of individuals arbitrarily or to his convenience.
- ❖ It constitutes the quota sample.
- ❖ It ensures that specific elements are included and represented in subsequent collection of them.
- ❖ The elements from the population are chosen on a non-random basis.
- ❖ All the members of the population do not have an equal chance of being selected to be a part of the sample group.

Advantages of quota sampling

- (a) It is useful when quick and crude results are desired.
- (b) It is convenient and less costly than other techniques.
- (c) It is independent of the existence of sampling frame.
- (d) When sampling frame is not available, it is perhaps the only choice available.
- (e) It gives guarantee the inclusion of individuals from different strata of population.

Disadvantages of quota sampling

- (a) Selected samples remain no longer representative of the population.
- (b) There is lack of external validity or generalization.
- (c) The researcher gets sufficient opportunity to select the most accessible individual (e.g. friends and relatives etc.), these accessible individuals may not be typical of the population of study.
- (d) It is amenable/ responsible to classification error.
- (e) It is less dependable and reliable.
- (f) The researcher controls one variable, but he/she cannot control other variable that may have both theoretical and practical significance.

(B) Convenience or accidental or incidental sampling

- ❖ It is also a non-probability sampling technique and also known as incidental sampling.
- ❖ The researcher selects the individuals according to his convenience.
- ❖ He is mainly focus on convenience and economy.
- ❖ He does not include the people with some specific trait.

Advantages of accidental or incidental sampling

- (a) It is most convenient technique of sampling.
- (b) It saves time, money and labour of researcher.

Disadvantages of accidental or incidental sampling

- (a) Samples remain no longer representative of the population. So, it cannot be done generalization.

- (b) The researcher gets sufficient opportunity to show his bias and prejudice in selecting individuals.
- (c) Probability of sampling error is high.
- (d) It is less dependable and reliable.

(C) Judgemental or purposive sampling

- ❖ It is based on the typicality of the case to be included in the sample.
- ❖ The researcher has some assumption that sample is very good representative of the population.

Advantages of judgemental or purposive sampling

- (a) It is very convenient technique of sampling.
- (b) It is less costly and more accessible to the researcher.
- (c) It guarantees that those individuals will be included in the sample that is relevant to the research design.

Disadvantages of judgemental or purposive sampling

- (a) There is no way to ensure that the samples are representative of the population. So, it cannot be done generalization.
- (b) The ability of the researcher is much emphasized in assessing which elements or individuals are typical population and which are not.
- (c) There is more subjectivity in sample selection.
- (d) In this type of technique, inferential statistics can't be used legitimately.

(D) Snowball or chain sampling

- ❖ It is a recruitment technique in which research participants are asked to assist researcher in identifying other potential subjects.
- ❖ When the researcher seeks to study a "hidden" population and the target population/characteristics are not easily accessible, e.g. drug addicts, hardened criminals, prostitutes etc.
- ❖ In such cases, this type of sampling technique is used.
- ❖ First of all, the researcher recognizes one potential subject, with the help of him he identifies the name and address of other subjects and these new subjects are told to give other names of subjects so that the growth of the sample tends to snowball. It revealed the pattern of diffusion of information.

Advantages of snowball sampling

- (a) It is a sociometric sampling technique. It has proved very important and helpful in studying small informal social group and its impact upon formal organizational structure.
- (b) It reveals communication pattern in community organization concepts.
- (c) It is responsible to various scientific sampling procedures at various stages, e.g. use of computer determination, use of random numbers etc.
- (d) It is important to apply in relatively small business and industrial organizations where N (population size) is not large or more than 100.

- (e) It is low cost.
- (f) It is useful in specific or rare circumstances.

Disadvantages of snowball sampling

- (a) It becomes heavy and difficult when N (population size) is large or more than 100.
- (b) The elements included in sample are not randomly selected and they are dependent on the subjective choices of originally selected respondents. It may introduce some bias in the sampling.
- (c) Bias can be presented in this sampling technique.
- (d) It has limited generalization.

(E) Saturation sampling

- ❖ It is a technique of selecting all elements/individuals having characteristics of interest to the researcher.
- ❖ It is when researcher reaches a point where no new information is obtained from further data.
- ❖ It indicates that adequate data has been collected for a detailed analysis.

(F) Dense sampling

- ❖ It is a technique of selecting elements/individuals between simple random sampling and saturation sampling.
- ❖ When the researcher selects his interesting 50% or more from the population with majority of individuals having specific traits or characteristics which are of interest to him, is known as dense sampling.

4. Data Collection

4.1 Concepts:

Data- Data is a plural number; its singular number is datum. It is any information or facts, figures, evidence, general material, or knowledge that is collected, observed, generated or created during the process of research study. These are observable and measurable facts that provide information about phenomenon under study.

Data collection- Data collection is the process of gathering and measuring information on variables of interest. It is an established system that enables one to answer stated research questions, test hypotheses and evaluate outcomes.

Data collection methods- The various strategies or steps used in gathering the research related data is called, data collection methods.

Data sources- Generally sources of research data are classified in two broad categories, primary and secondary data sources. Primary data are directly collected from primary sources (individuals, objects, institution etc.). They (primary sources/ respondents) provide the first-hand information/data to researcher directly. These data may be collected through various research tools/ techniques (interviews, questioning, observation, biochemical measurement and psycho-social measurement scales). Secondary data are directly collected from external and internal secondary sources. The external sources are published records (journals, magazines, newspapers, government

reports, census reports commission reports etc.) and published records (official records, patient records, thesis, dissertations etc.). The internal secondary sources/private documents are (biographies, personal diaries, letters, memories etc.).

Data collection plan- It is data collection strategies, which would produce meaningful factual information of data and which may answer the research question. It includes what data is to be collected, how the data is to be collected and what will be the data characteristics. It determines the following identifications:

(i) Data needs

Identifications of data needs are done for addressing the research questions, testing hypotheses, describing the main characteristics of the sample (age, gender, race, religion, ethnicity, education level, occupation, income and economic status, family status, social status, marital status and other relevant information depending on the type and topic of the research study). The information related to population characteristics is helpful in interpretation and generalization of the research findings.

- ❖ Identifying sources and isolating extraneous variables from key variables.
- ❖ Selecting types and methods of measurement of all the variables.
- ❖ Ensuring accuracy in data measurements.
- ❖ Analyzing potential biases and make plan for removal of biasness.
- ❖ Accuracy in statistical findings.

(ii) Data characteristics

It has following dimensions:

Data structure- Data may be two types: structured data and unstructured data. Structured data are collected through a structured research tool- where the questions are asked in a specified sequence and the respondents are directed to select the "pre-determined answers/options." These are also convenient to analyze. Unstructured data are those, where the respondents can answer openly and freely in their words/language. These are difficult to analyze.

Data quantifiability- The numerical values of data are measurable or quantifiable, it is easily used in statistical computation. Structured data are more useful/ convenient in statistical computation than unstructured data.

Data objectivity- It means data is free from the researcher's personal biases. The research findings of data collected by two researchers in same study independently and separately will match, if there is objectivity in data.

Data obstructiveness- During data collection process, it is necessary for the researcher to make an effort to put the respondents at ease. Make them familiar with research topic and make them comfortable so that they may give their natural responses/ behaviours.

(iii) Data sources

The researcher needs to identify the sources of data; it may minimize the time and the cost for data collection. These may be classified as below:

Primary data sources- The researches published by the researcher who conducted it is called primary data source. Primary data are directly collected from primary sources (individuals, objects, institution etc.). They (primary sources/ respondents) provide the first-hand information/data to researcher directly. These data may be collected through various research tools/techniques (interviews, questioning, observation, biochemical measurement and psycho-social measurement scales).

Secondary data sources- Secondary data are directly collected from external and internal secondary sources. The external sources are published records (journals, magazines, newspapers, government reports, census reports commission reports etc.) and published records (official records, patient records, thesis, dissertations etc.).

Personal data sources- These includes the live data sources (such source may be a person, who gives a personal account of the experiences, events etc.), e.g. an elderly freedom fighter of pre-independent era of India, who has personally gone through various political events of historical period, who suffered the pre-independent struggle of the country during the British colonial rule. Biographies, personal diaries, letters, memories etc. are included in the secondary sources/private documents.

Historical data sources- These includes all the historical documents, letters, coins, diaries, autobiographies of historians, artifact etc.

Documentary data sources- These includes all the documented records and reports (patient care records, medical records, academic records, service records, financial records and reports, government and international reports)

(iv) Categories of data

The researcher needs to categorize the data. These may be classified as below:

Personal data- It is also called identification data which reveals personal information or data about the respondents. It includes the respondent's name, age, gender, race, religion, ethnicity, education level, occupation, income and economic status, family status, social status, marital status etc. and other relevant information depending on the type and topic of the research study).

Factual data- It includes detail factual information of any organization, institution, any event and specific activities related to country planning, implementation and outcome.

Data on attitude and feelings- It belongs to the affective domain which describes about attitude and feelings of a target population. It may be subjective data as one's personal attitude and feeling as expressed towards some subjects. The data may also be objective data if collected through standardized tools.

Data on judgements and options- It involves factual information or data about their version of judgements and options based on logic, reasoning, justification etc. The level of judgements is measured regarding one's personality, behaviour, performance etc. These judgements and options may be on persons, professions, poli-

cies and programmes etc.

Data on psycho-motor skills- It involves the data which is collected through event observation of a specific skill (physical and mental).

Data on achievement and accomplishments- It involves the data of an individual's performance in his/her achievement in academics/sports/music etc. The achievement may be social, political, economic etc. These achievement tests are used for grading, promotion, fellowship, scholarship etc.

Let us check our progress-4

Write the meaning of followings:

(a) Data sources

(b) Categories of data collection

4.2 Data collection techniques

It is a process, in which the researcher chooses an appropriate data collection approach for gathering reliable, accurate and valid data. Data collection is the most time-consuming and costly part of research. While selecting the data collection techniques, issues related to ethics, time, cost and manpower should be considered.

Qualities of good data collection techniques:

(A) Reliability- It should be reliable. It is the degree to which a measuring tool/instrument gives the same results each time that it is used.

(B) Validity- It should be valid. It is the degree to which the tool/instrument measures, what it is intended to measure.

(C) Objectivity- It should be objective. It means the technique should free from personal biases.

(D) Accuracy- It should be accurate. It means the technique should free from all errors.

(E) Simplicity- It should be simple. It means the technique can be operated in a simple and easy manner.

(F) Economic- It should be economic. It means the technique should have quality of time and cost effectiveness.

Classification of data collection techniques

Following are the main data collection techniques:

(A) Questioning- It means "the action of asking someone question." It is a technique for data collection in research process, where series of questions are asked to elicit factual information from the respondents related to some particular event or issue. It is presumed that the respondents have knowledge on it. The answers of the questions may be asked verbally as well as written.

Questions in questioning technique may be classified in following types:

(i) Closed ended or structured questions- These are a set of questions. Answers of these questions are predetermined as options. The respondents are directed to choose most suitable answer from the given options against each question. It is dif-

difficult to construct these types of questions but respondents find easy to answer. The answers are found ready-made; respondents have only to choose from the given options. Those respondents, who are not convenient to express writing the answers, they find it suitable. Data obtained through these types of questionnaires are easy in organization and analysis. Closed ended or structured questions may be further classified in following types:

- (a) **Dichotomous questions-** These questions contain only two options as answer. The respondents are to choose one most suitable between them.
- (b) **Multiple choice questions-** These questions contain a range of 3-5 options as answer. The respondents are to choose one most suitable among them.
- (c) **Rank order questions-** These questions contain answers which are ranked on a continuum. The respondents are asked to place the answer options on ranks according to their choice.
- (d) **Rating questions-** These questions contain answers options on a scale which have number of gradations or points along the scale in continuum having two opposite extreme situations at both ends.
- (e) **Matrix questions or checklist-** These matrix questions or check-list questions are two-dimension arrangement of several questions, the respondents are directed to answer in the same format. The questions are presented in vertical series and the answers are given in horizontal series.
- (f) **Calendar questions-** These questions are used when there is need of data on chronology of events. Questions are related to beginning, duration and end of the events.

Advantages of closed ended or structured questions

- (a) It is convenient for those respondents who are not very expressive.
- (b) A large number of questions can be included in it.
- (c) It is not time-consuming.
- (d) Obtained data are highly organized, so it is easy to analyze.
- (e) The researcher can observe each and every step of the occurrence of the events.
- (f) The researcher's directions to the respondents for purpose data can be maintained.
- (g) It prevents subjectivity in answering.
- (h) It is easy to do statistical calculations.

Disadvantages of closed ended or structured questions

- (a) It is difficult to construct closed-ended questions.
- (b) Obtained data are superficial in nature.
- (c) The options of answers for the questions are fixed. The respondents are not free to think other answers.

(ii) **Open-ended or unstructured questions-** These are also a set of questions.

Answers of these questions are not predetermined as options by the researcher. The respondents are directed to answer freely, spontaneously in their own words and own language. It plays a key role where research is done in new area because these questions allow more depth thinking related to research issues. The respondents do not face any restraints, so these data are more reliable. Those respondents, who are very expressive in giving the answers, they find it suitable. These are usually used in qualitative research. The researcher needs experience and skills to construct these types of questions. Data obtained through these types of questionnaires are difficult in organization and analysis.

Advantages of open-ended or unstructured questions

- (a) It is convenient for those respondents who are very expressive.
- (b) It is flexible in nature.
- (c) The respondents elaborate the responses.
- (d) Factual information can be collected in detail.

Disadvantages of open-ended or unstructured questions

- (a) The options of answers for the questions are not fixed; the respondents are free to think other answers. During thinking and elaboration of the responses, the respondents may get diverted from the main objectives.
- (b) The researcher finds himself unable to get the correct and required answers.
- (c) Obtained data are not highly organized, so it is difficult to analyze.

(iii) Pictorial questions- The respondents are provided a set of pictures. A set of questions is also given, which is based or closely related on the pictures. The respondents are asked to answers to the questions based on a set of pictures. The researcher collects the information, organize and analyze it.

Classification of questioning techniques approach

Questioning techniques approach may be classified as follows:

(i) Self-reporting approach- It is a data collection technique, in which questionnaire of a series of pre-set questions are given to respondents. The respondents are directed to answers to these questions in writing. It is also called paper-pencil approach. It can cover a wide variety of issues. Such data have doubt in the validity and accuracy, because the respondents may not give frank or true information.

Advantages of self-reporting approach

- (a) It can cover a widely scattered area.
- (b) It provides freeness to respond.
- (c) It is economical in terms of time and cost.
- (d) It can use standardized questionnaire.
- (e) There is possibility to maintain full uniformity and objectivity.

Disadvantages of self-reporting approach

- (a) If the question is not clear, respondents may not reply all the questions as directed.
- (b) If data are collected through mailed questionnaire, some respondents may not return.

(c) The researcher is not able to judge the respondents' reaction to the question.

(ii) Interview approach-It is a data collection technique, in which information or data are collected through face-to-face meeting or through telephonic interaction. In this technique the researcher asks verbal questions and the respondents reply to the questions. The researcher either notes the verbal answers or records with the help of tape-recorder. It is a more personal interaction between the researcher and the respondents. It may be for a single person (personal interview) or for a group (group interview). When it is taken for a group of 5-10 respondents, it is called focus group interview.

The researcher is able to connect all the group members at the same time through a single interview. But of the group members may not feel comfortable to express themselves in presence of others.

Advantages of interview approach

(a) It gives chances of more personal interactions.

(b) It is fit for those respondents who prefer verbal interview in spite of written.

Disadvantages of interview approach

(a) It is more time-consuming and expensive.

(b) The researcher needs to arrange a venue for interview.

(c) The researcher's personal biases may affect data.

(d) It may have lack of objectivity and reliability.

Types of interview

Types of interview may be classified as follows:

(i) Structured interview- In this type of interview, the researcher prepares a set of questions on a desired topic or issue. The interview questions are set, concrete, fixed. It is scheduled for a fix duration in which the researcher asks questions face-to-face. The responses are recorded through notes taking or with the help of tape-recorder. It is also done through telephonic contact. The respondents-researcher conversation is taken place on telephone. The conversation is recorded automatically in system. The researcher guides and helps in focusing towards the goal. At the end, interviewer concludes the interview following a vote of thanks.

Advantages of structured interview

(a) The data collected through structured interview are highly organized.

(b) It is easy to analyze the data.

Disadvantages of structured interview

(a) The questions for interview are fixed. During process of interview these questions cannot be changed.

(ii) Semi-structured interview- In this type of interview, the researcher prepares a set of questions on a desired topic or issue. The researcher asks questions face-to-face during interview schedule. He or she encourages the respondents to express freely about their feelings, opinions and perceptions. The obtained data are scattered

so makes difficult to analyze. The responses are recorded through notes taking or with the help of tape-recorder.

(iii) Unstructured interview- In this type of interview, the researcher does not prepare any set of questions on a desired topic or issue. The interview questions are not set and fixed. The researcher asks non-specific and general questions. The questions are asked regarding a desired topic or issue. The respondents are allowed to take part freely, frankly and truthfully. The researcher is free to frame questions according to needs. It is playing a key role if the research is conducting in new area with new problem. The responses are recorded through notes taking or with the help of tape-recorder. The researcher guides and help in focusing towards the goal. At the end, interviewer concludes the interview following a vote of thanks.

Advantages of unstructured interview

- (a) It gives chances of more personal interactions.
- (b) It is fit for those respondents who prefer verbal interview in spite of written.

Disadvantages of unstructured interview

- (a) It is more time-consuming and expensive.
- (b) The researcher needs to arrange a venue for interview.
- (c) The researcher's personal biases may affect data.
- (d) It may have lack of objectivity and reliability.

Steps of interview

- (a) Preparing an interview schedule with date, time and venue.
- (b) Preparing a list of questions which are to be asked in the interview.
- (c) Selecting the respondents/interviewees for the interview.
- (d) Informing the respondents/ interviewees about the interview date, time and venue.
- (e) Arranging the venue of interview- lighting, comfortable seating accommodation prior to date of interview.
- (f) On the day of interview, the interviewer/ researcher should reach before interviewees.
- (g) Receiving cordially, introduces self, lead them to the interview hall and make them seated the respondents/interviewees at the venue of the interview by the interviewer.
- (h) Suggesting respondents/ interviewees to introduce themselves to each other. It will help in rapport building and ease feeling among group members.
- (i) Discussing briefly the objectives of the interview and research topic.
- (j) Assuring the respondents/interviewees about the confidently of their identities and contents of the discussions.
- (k) Starting the interview with a few general questions and then come to the point/question list (if, it is structured interview).
- (l) Asking the questions one by one (if, it is group interview).

- (m) Encouraging the respondents/interviewees to speak freely, frankly and elaborately.
- (n) Making sure that all the participants are involved in question-answer session.
- (o) Listening carefully and attentively when the respondents/interviewees start speaking.
- (p) Making the appropriate notes on the paper or record with the help of tape-recorder. Do not miss any part or change in words of the respondents/interviewees.
- (q) Clarify, if the respondents/interviewees are finding difficulties in understanding the questions.
- (r) When researcher feels that answers are inadequate, encourage to explain details.
- (s) Making sure that the discussions or responses are relevant to the questions.
- (t) Conveying sincere thanks to the respondents/interviewees for their participation and co-operation and end the interview.
- (u) Serving tea and refreshment to the respondents/interviewees.

Qualities of an interviewer as researcher- An interviewer as researcher should have following qualities:

- (i) An interviewer as researcher should be a good questioner.
- (ii) He/she should be able to overcome the communication barrier by speaking the same vocabulary and language according to respondents.
- (iii) He/she should be an intent listener so he/she can prepare appropriate follow up questions (if, it is unstructured interview) and able to record accurate notes of responses.
- (iv) On the day of interview, the interviewer/ researcher should reach before interviewees and meet with the respondents to get familiar to obtain maximum co-operation from the respondents.
- (v) He/she should ensure objectivity and reliability in recording of data.

(B) Observation- It is data collection technique and used where the important or critical events are observed visually. It is widely used to observe the respondents' behaviour or their reactions to a particular critical event/situation. It is used in variety of situations. The purpose of this technique is to observe the sequence of events step by step. It is useful in following situations-

- (i) When the respondents are not aware of their behaviours or activities- In such conditions the researcher can observe their spontaneous behaviours, activities or reactions.
- (ii) When the respondents are not able to answer questions- In such conditions the researcher can observe their spontaneous behaviours, activities or reactions, e.g. patients in coma state or small children etc.
- (iii) When the respondents are not able to articulate their speech- In such conditions the researcher can observe their spontaneous behaviours, activities or reactions,

e.g. mentally ill patients, small children etc.

- (iv) When the respondents are under the grip of severe emotional stress, grief or mental shock.
- (v) When the respondents feel embarrassed to report their activities or behaviour.

It may be classified as follows:

(i) Visibility

(a) Direct observation- The researcher observes the respondents' behaviour or event directly and personally.

(b) Indirect observation- The researcher observes the respondents' behaviour or event indirectly and with electronic equipment indirectly and with an electronic equipment e.g. video camera or live telecast on the television.

(ii) Continuity

(a) Constant or continuous observation- The researcher observes the respondents' behaviour or event continuously, e.g. observation of operation procedure in operation theatre.

(b) Intermittent observation- The researcher observes the respondents' behaviour or event intermittently, e.g. an infants' growth is observed once a week etc.

(c) Instantaneous observation- The researcher observes the event instantly when there is sudden or unpredictable happening, e.g. an earth quake, an accident, a fire breaks out etc.

(iii) Participation

(a) Participant observation- The researcher observes the respondents' behaviour or event as active participant. The disadvantage of this type of observation is that the researcher may miss items of observation or the observation may get biased.

(b) Non-participant observation- The researcher observes the respondents' behaviour or event not as a participant. He functions as observer only. After observation is over the researcher prepares a written report of the observation.

(iv) Appearance

(a) Concealed observation- Concealed observation is a research method where the observer remains hidden from the subjects of the study. The researcher will not reveal their research purpose in concealed observation.

(b) Revealed or non-Concealed observation- It is a study method in which the observer is not concealed from the test subjects. In revealed or non-concealed observation the "participants" are aware of the research purposes.

Advantages of observation

- (i) The researcher observes freely, he does not depend on the responses of respondents.
- (ii) All the subjects in the sample under observation.
- (iii) According to need, researcher can observe continuously by a video camera or

he can observe intermittently by a plain ordinary camera.

- (iv) Observation done by a video camera or a plain ordinary camera ensures objectivity and accuracy.
- (v) The researcher can observe each and every step of the occurrence of the events.
- (vi) The researcher is free to start or stop observation anytime.
- (vii) It is more reliable and free from biases.

Disadvantages of observation

- (i) It is time consuming technique. The researcher has to wait for the occurrence of event.
- (ii) The researcher's observation may be biased because of his subjectivity.
- (iii) The interpretation of observed events may not be proper and accurate because it depends on researcher's academic and cultural background.
- (iv) The interpretation may be incorrect, if the researcher is not skilled observer or interpreter.
- (v) It is time consuming technique because the amount of data gathered is very large, it needs long time to organize and analyze the data.
- (vi) It is very difficult for the researcher to observe and record the event personally and simultaneously without any visual aids.
- (vii) The respondents may change their originality and spontaneity of their responses when they aware that some is observing them.
- (vii) This technique needs extensive trained, efficient and skilled observer. Sometimes this type of observers are not available, so quality observation and interpretation cannot be possible. It increases biasness.

Observational biases

It may occur due to following reasons:

- (i) Observational process may be affected by improper or inadequate skill of the researcher.
- (ii) The researcher's personal interest may deviate the direction of observation.
- (iii) The researcher's expectation of the observational findings may lead to erroneous observation.
- (iv) Personal attitude, values, prejudices may influence the interpretation of the observational findings and may result into faulty inferences.

(C) Measurement- The factual data are obtained with the help of various standardized measurements. It can be quantitative or qualitative measurement. Quantitative measurement is a process of ascribing numerical values/scores to objects or events which represent the amount of their characteristics. Qualitative measurement is a process of classification which assigns qualities to certain categories.

Measurements may be classified as follows:

- (i) **Physical measurements-** These are related to people's height, weight, body circumstances, skin fold thickness and body surface area etc. These are

obtained by various measuring tools/equipment e.g. weighing machine, height scale, tape measure etc.

- (ii) **Physiological measurements-** These are related to various body functions and includes haemoglobin level, blood sugar level, blood urea level, blood cell count, body temperature, pulse rate, respiration rate, heart rate, blood pressure etc. These are obtained by various measuring tools/ equipment e.g. thermometer, blood pressure apparatus (sphygmomanometer), cardiac monitor and laboratory machines etc.
- (iii) **Psychological measurements-** These are related to people's personality status, intelligence level, aptitude, attitude, feelings, anxiety level etc. These are conducted with the help of psychometric tests.
- (iv) **Sociological measurements-** These are related to people's social status, economic status etc. These are conducted with the help of socio-metric tests.
- (v) **Environmental measurements-** These are required when research studies are done related to environment e.g. climatic conditions (temperature, humidity, rainfall etc.) are measured.
- (D) **Record analysis-** It is used when there is need of some recorded data from the documents or records. It may be taken from government records, health records, academic records, service records, annual reports etc. It is most cost effective and less time consuming but it may be incomplete and inconsistent.
- (E) **Life-history analysis or autobiographies-** These are used in historical research studies. These are life-history or autobiographies of world fame leaders in areas of political, religious etc. The information obtain from these sources is unstructured in nature.
- (F) **Diaries and letter analysis-** These may provide data of both the historical and non-historical type. These may provide personal experiences of specific period related to research. The information obtain from these sources is also unstructured in nature.
- (G) **Critical incident analysis-** The researcher describes a critical incident like Ukraine-Russia war. It may be negative as Indian medical students faced the situation or positive when they success in escaping their life safe. The respondents are asked/ directed to recall their personal experiences about the similar incidents or situations to describe. The researcher observes the respondents' reactions while describing about critical situation. It focuses on the factual information that may be elicited from their describing and reactions etc.
- (H) **Q-sort analysis-** The researcher makes some descriptive statements about the related incidents/events. The respondents are asked/directed to arrange these statements according to degree of their likes and dislikes. It is presumed that they have knowledge about it. It is useful in research on people's behavioural characteristics.
- (I) **Instance analysis-** The respondents is asked to write down experience about the

related incidents/events. The respondents write their experiences from their memory. The written report is matched with factual data.

Let us check our progress-5

Describe data collection techniques

4.3 Data collection tools

After deciding or selecting the data collection technique by the researcher, next step is to find out/select or construct the data collection tool suitable to the research study. Data collection tool/instrument is a device/equipment, which is designed or developed carefully by the researcher. It contains questions related to the research topic under study.

Purpose of the data collection tool

- (i) To establish a set of criteria upon which the research conclusion is to be established.
- (ii) To provide measurement of selected criteria.

After deciding the data collection tool, the researcher needs to do the following things:

- (a) The researcher should find out whether the required tool is already available.
- (b) Whether the required tool is already available, the researcher should review the tool for its appropriateness.
- (c) The researcher should need to determine, is it conceptually relevant.
- (d) When the required tool is already available and appropriate for fulfilling the purpose of data collection for research.
- (e) The researcher should take permission from the concerned authority to use the tool.
- (f) When the required tool is already available, but needs some modification to serve the purpose, then the permission is required from the concerned authority.
- (g) When the required tool is already not available, then the researcher should construct or develop a tool for fulfilling the purpose of data collection for research.
- (h) When the required tool is constructed or developed, it should be examined by the language expert.
- (i) After language examination, it should be sent to the research expert for its clarity, objectivity, reliability and validity.
- (j) After finalized the data collection tool, it should be pre-tested through pilot study to evaluate the appropriateness and refine the prepared tool.
- (k) The researcher does the required modification and corrections to make ready the tool for the final data collection.

Pre-testing of the data collection tool- It is the process of administering the tool to a similar group or sample obtained from the target population. It is done to test the reliability and the validity of the tool. The purposes of the pre-testing tool are following:

- (i) To find out time requirement for administering the tool.
- (ii) To find out difficulty in any item or question to understand.
- (iii) To identify inappropriateness, offensive or objectionable items or questions, this

requires deletion or omission.

- (iv) To determine the validity, objectivity, specificity, sensitivity, accuracy, precision, reliability and the generalization of the tool.

General characteristics of a data collection tool

The data collection tool should have following general characteristics:

- (a) It should be relevant to the research topic.
- (b) It should support the research approach.
- (c) It should be based on the theoretical framework.
- (d) It should be attractive and pleasant in appearance to work.
- (e) It should explain the purpose of the research study.
- (f) It should contain clear instructions about use of tool, to the respondents.
- (g) It should contain legible questions, so that the respondents can answer conveniently and completely.
- (h) It should have questions in understandable language.
- (i) It should have quality of reliability, validity, objectivity, generability etc.
- (j) It should maintain the ethical standards to preserve privacy and anonymity of the respondents.
- (k) It should maintain the identities and confidentiality of responses of the respondents.
- (l) It should be free from any personal biases.

Qualities of a data collection tool

The data collection tool should have following qualities:

(i) **Validity-** The data collection tool should valid. It means, it has a capacity to measure what it actually tends to measure or it is the degree to which the tool/instrument measures, what it is intended to measure. Validity may be divided as follows:

- (a) **Content validity-** It refers to the items in the tool, if they are systematically organized, the items cover enquiry in a balanced way, items express the scope of the research study and items reflect the concept under consideration.
- (b) **Construct validity-** It focuses on the abstract concept and its relationship to other concepts.
- (c) **Criterion and related validity-** It represents the associations between the two measures in the same phenomena. It compares well with the external variables considered to be the direct measure of the behaviour or performance in the actual situation. It is further divided into-
 - (i) **Concurrent validity-** It relates to the performance at approximately the same time when the data collection tool is administered.
 - (ii) **Predictive validity-** It refers to the future performance or forecasting of future criterion of the tool.
- (d) **Logical validity-** It refers to that aspect of a tool to find out if the tool collects data, what it is supposed to collect and to what extent of specification.

- (e) **Empirical validity-** It refers to the effectiveness of the tool in predicting successful performance according to its design or construct.
- (f) **Face validity-** It refers to the exterior or appearance of the tool with an expectation that it will function as an effective one.
- (ii) **Reliability-**The data collection tool should be reliable. It means, whoever and whenever the tool is administered, it will elicit same data. Reliability refers to the in-built excellence and integrity of the data collection tool. It is the degree to which a measuring tool/instrument gives the same results each time that it is used.
- (iii) **Objectivity-**The data collection tool should have objectivity.
- (iv) **Accuracy-**The data collection tool should have accuracy. It refers to the degree to which the tool obtains true data without any error.
- (v) **Precision-** The data collection tool should have precision. It is an estimate of exactness with which the tool performs its function which can be repeated and which is free from all variation.
- (vi) **Specificity-** It refers to that aspect of the tool functioning, which brings perfect results in similar circumstances.
- (vii) **Sensitivity-** The data collection tool should be sensitive. It refers to what aspect of the tool where the tool is able to consistently discriminate even smallest differences between two samples and the smallest amount of variation in a given range of error.
- (viii) **Generalizability-** It refers to the degree to which data collection can be generalized.

Possible sources of errors or inaccuracy

- (i) Errors in the tool.
- (ii) Errors in the settings or methods.
- (iii) Inconsistency in terms of administering the tool.
- (iv) Inconsistency in terms of subjects to whom the tool is administered.

Types of data collection tools

Data collection tools are classified as follows:

(i) **Questionnaire-** It is most frequently used tool. In a questionnaire, a set of questions are arranged in an orderly sequence. It is used to collect the factual data on certain population, organization, community etc. The administration of questionnaire is convenient; it can be delivered by hand or by mail. A population of a wide geographical area can be covered through this tool.

Essential characteristics of a questionnaire as data collection tool

- (a) It should contain purpose of the research, so that the respondents may get general idea about the research topic.
- (b) It should have questions related to the research topic.
- (c) It should contain the instruction to respondents about how to answer and the dead line of return of the questionnaire.
- (d) It should have a covering letter, a self-addressed stamped envelope for returning of the questionnaire.

- (e) It should have short, simple and in understandable language.
- (f) It should contain questions free from subjectivity and ambiguity.
- (g) It must avoid double-barreled or negative questions.
- (h) It should have a movement of question from general to specific.
- (i) The responses should not have two opposite ideas.
- (j) It should be provided adequate choices in closed-ended questions.
- (k) It should maintain the ethical standards to preserve privacy and anonymity of the respondents.
- (l) It should maintain the identities and confidentiality of responses of the respondents.
- (m) It should have quality of reliability, validity, objectivity, generability etc.

Advantages of a questionnaire as data collection tool

- (i) It is more cost-effective and time-effective.
- (ii) It covers a large number of populations.
- (iii) It can maintain the respondents' anonymity.
- (iv) It provides highly organized data with closed-ended questions, are easy to tabulate and organize.
- (v) It provides a large amount of data which are used in quantitative research studies.
- (vi) Close-ended questions can reduce personal biases on the part of the researcher.

Disadvantages of a questionnaire as data collection tool

- (i) It is difficult to construct.
- (ii) If, there are some difficult or unclear questions that may not be completely answered.
- (iii) Some respondents do not return questionnaire completely or timely.
- (iv) Certain groups of respondents (small children, illiterate population etc.) cannot be covered by this tool.

(ii) Opinionnaire- It is used for collection of data on respondents' opinions on some issue related to the research topic. It contains a set of opinion-statements, needs respondents' opinions. The opinions are people's beliefs or feelings. These are closely related to the respondent's measurable behavior. It makes use of checklist which comprises a series of statements. The statements are descriptions or traits of individuals, which the researcher checks as being present or absent/ yes or no in individuals. The type or number of items may be indicated by ascribing appropriate grade or score.

(iii) Measurement scale- It is a data collection tool which assign numerical values or scores for continuing measuring attributes. It is a set of symbols or scores so that these symbols or scores can be ascribed to various characteristics of the subjects, for whom the scale has been used. It measures these characteristics which place each subject on a continuum.

The measurement scales are as follows:

(a) Rating scale- It is a scale, where numerical values are attributed to a series of items or statements designed to assess the subjects' feelings towards special phenomena (characteristics, places, situations, people, policies etc.) with varying degree of intensity (high, medium, low etc.) in a scale that will reflect their likes, dislikes etc. It involves qualitative description of characteristics of a person or a thing. It is useful in making and recording observation. The data provided by these scales can be statistically treated; the reliability of the rating scale is very high. In these scales, the classification may set up in 5-7 categories. It is easily constructed but difficult to define clearly the traits or the attributes.

(b) Likert scale- It is constructed after name of a psychologist, Rensis Likert. It is a scale with five or more assessment statements to which numerical values or scores are ascribed. These points state are opinions on a topic. The highest score is ascribed to the most agreeable (positive) statement and the lowest score is ascribed to the most disagreeable (negative) statement. It is used to measure attitude of the respondents on a specific issue. The respondents are directed to express their agreement with the stated viewpoints.

Construction of a Likert scale- Following steps may be adopted:

- (i) Collect a large pool of statements which reflect individuals' attitude towards topic.
- (ii) To avoid biases, equal number of positive and negative statements is selected. These are usually selected 5-7 categories of positive as well as negative statements on a 5-7 points scale.
- (iii) A favourable view point is assigned to the highest score so that, high score will consistently reflect positive view points.
- (iv) Analyse the items to identify their discriminating values and retaining items which have large discriminating values e.g. highly positive or highly negative and finally include these items in the final scale. After each item is scored, a total score is computed.

Advantages

- (i) It is simple to construct.
- (ii) It is highly reliable.
- (iii) In this type of tool, the range of responses provides more precise and more effective information about the subjects' attitude.

Disadvantages

- (i) It does not specify about which item is more agreeable.
- (ii) Total score does not give a clear attitude.

(c) Semantic differential scale- It is a measuring tool, used to measure rate of the individuals' responses about a particular concept on a bipolar adjectives-strong/weak, good/bad etc. A seven-point rating scale is used where the line moves from one extreme dimension to the other extreme. The scores are then summed up to get a total score. It is easy to construct but it needs to pay attention while selecting appropriate adjective-

pairs. Following adjectives are usually used:

- (i) Evaluative adjectives- Good/bad, effective/ineffective
- (ii) Potency adjectives- Strong/weak, large/small
- (iii) Activity adjectives- Active/passive, fast/slow

(d) Visual Analogue scale- It is used to measure subjective experiences or feelings of the individuals, e.g. pain, anxiety, weakness etc. It is either on vertical line or horizontal line, have various points on a straight line. These points indicate various amount of experiences or feelings of the individuals. The beginning point of scale represents the lowest score or zero and the end point of scale represents the highest score, e.g.100. It is used to collect interval data where the subject is directed to indicate a point on the linear scale that reflects the intensity of their experiences.

Let us check our progress-6

Discuss data collection tools.

4.4 Pilot Study

The pilot study is done to find out the feasibility of the study, test the tool, select the appropriate methodology and identify the potential problems in research study. It helps the researcher to acquire knowledge about the target population, validity, reliability and effectiveness of data collection techniques and tool. It also helps in improvement of the various aspects of the main research study. It is a rehearsal study or the whole study in miniature.

Purposes of the pilot study

Following are the purposes of the pilot study:

- (i) To test the data collection technique to the target population but not to the same subjects, those who will participate in the main research study.
- (ii) To test data collection tool its effectiveness and efficiency (its accuracy, reliability, validity etc.) if, data collection tool needs some modification or correction.
- (iii) To detect any hidden weakness or errors in the research process.
- (iv) To estimate time and possible cost.
- (v) To check all the phases of research process beginning from to end.

Let us check our progress-7

Write meaning of a pilot study.

4.5 Data Collection Procedures

After completion of pilot study, the next step is the stage of implementation. Before we actually go for data collection, it is necessary to make a plan and to find out

(i) What data are to be collected?

The researcher needs to consult statistician about the type or form of data that would get after data collection. It should make sure that data collected would match to the analysis plan. Because, some statistical test can only be done on particular data, e.g. interval or ratio data.

(ii) From whom are the data to be collected?

It involves the consideration of sample subjects (sample, their selection procedures, numbers etc.). Usually a detail plan is required which indicates the nature of the study, subjects involved, kind of independent variables. The permissions/ ethical clearance from the authorities are to be taken. It protects the researcher as well as subjects.

(iii) Who will collect the data?

The researcher and his team are responsible for data collection. A pre-test is needed; mostly it is carried out during the pilot phase of the study. If the researcher appointed some people for data collection, they should be trained thoroughly, so they can collect data accurately and completely.

(iv) From where will the data be collected?

According to research problem, it needs to collect data about the location, communication, climate, language spoken, food habits, religion etc. In an experimental research, it should not be mixed the control group and experimental group subjects.

(v) When is the data to be collected?

The researcher should have knowledge about the geography and society of the population of research. So that he can handle all the circumstances during the data collection.

Data collection through questionnaire

The researcher should contact and build rapport with the subjects, to be included in research study. At the time of questionnaire delivery (by hand or by post/mail), it is necessary to instruct the respondents about the way of answering the questions and time of return back. A self-addressed envelope with postal stamps and follow up letter of request also be attach with questionnaire. Finally, obtained data are recorded and kept ready for tabulation and analysis.

Data collection through interview

The researcher prepares interview schedule and the list of questions for the interview. These are prepared in form of structured or unstructured interview. The researcher informs the schedule to all the participants/respondents. During interview session, all the participants/respondents are made relaxed and comfortable. The data obtained should be noted carefully or recorded with the help of tape-recorder.

Data collection through observation

The researcher should prepare tool which will guide and ensure systematic observation. The researcher informs the observation schedule to all the participants/respondents. Observation is done according to decided observation technique and data are recorded accordingly. Data may be recorded with the help of ordinary camera or video camera.

Data collection through case studies

In case studies, data are collected from an individual, a family, an institution or a

community. After collection of data, editing and classification is done. Data editing includes removing all deficiencies (data errors, data incompleteness, data inconsistency etc.). It should be ensured that all the data should follow and maintain uniformity and consistency. Data classification is the process of logical dividing of data in various categories. The collected heterogeneous data are transformed into some homogeneous form.

Let us check our progress-8

Narrate the data collection procedures.

- 5. Data Organization** - It is arranging the gathered data in a systematic way so that they appear meaningful. So, the data needs to be organized into groups. The qualitative data needs to be coded through a coding system which allows numerical analysis of a qualitative data. The scores of raw data are difficult to interpret, so the scores are to be arranged in descending order of magnitude which is called array. It provides a system of arrangement where it is easy to work on the data. The highest score, middle score and the lowest score can be identified with the help of array.
- 6. Data Presentation-** After organization data, it is presented in a meaningful form. It consists of data summarization (transforming the data into frequency distribution) and display of data. The summary presents the distribution of data in their numerical scores. The data is presented or displayed in the following manner:

(i) Data presented by table

It is a process in which the numerical data is put in a tabular form on a tabulation sheet manually or with the help of a computer. Following important points to be kept in mind while data are presenting a tabular form-

- (a) The table should be numbered by Arabic or Roman numbers.
- (b) There should be given a title of the table. The title should exactly depict what tables present.
- (c) The vertical presentation in table is called, columns whereas the horizontal presentation in table is called, rows.
- (d) The table should have simple, clear and understandable presentation of the data.
- (e) It should be used one table for one type of data.
- (f) Mostly, in the table, the numerical data is usually presented in descending order of magnitude of frequency.

(ii) Data presented by graph *(please see examples of figure of graphs on chapter 6 para 5.2)*

The data can be presented by graph. It presents relationship between various types of data. The graphs may be following two types-

- (a) Line graph-** Graphs are prepared with reference to two coordinate lines. The one vertical line or Y-axis and the other horizontal or X-axis. These two lines or axis are perpendicular to each other. The intersecting point of the two axes is called, zero or point of origin. It is the point of reference for both the axes. Graphs are presented on the horizontal axis from the left to the right and on the vertical axis from the

bottom to the top. The distance measured along the X-axis to the right of the point of origin is called, positive. The distance measured along the X-axis to the left of the point of origin is called, negative. The distance measured along the Y-axis above the point of origin is called, positive. The distance measured along the Y-axis below the point of origin is called, negative. Both the axis (x and y) make quadrants. The right upper quadrant (first quadrant) measures positive (++) on both the axis (x and y). The left upper quadrant (second quadrant) measures negative on X-axis and positive on Y-axis (- +). The left lower quadrant (third quadrant) measures both the axis (x and y) are negative (- -). The right lower quadrant (fourth quadrant) measures X is positive and the Y is negative (+ -).

Usually the horizontal axis (X-axis) of graph measures the independent variables whereas the vertical axis (Y-axis) of graph measures the dependent variables. The intervals in the graph should be equal.

The line graph is used in making predictions and also in presenting relationship between different types of data.

(b) Bar graph- In the Bar graphs, bars are separated by space. When the bar graphs need to contain many items, then the bars are joined to save space. It can be further divided into following two types:

Vertical bar graphs- These are used to compare all different times.

Horizontal bar graph- These are used to compare the components at a particular time.

(iii) Data presented by pie-chart

Pie-chart presents various components of the data in a segmented circle with simple percentage distribution. The radius is drawn vertically and the components are arranged in a clockwise direction in descending order of frequency or magnitude. The proportion of data is indicated by the number of degrees in each segment of the 360° circles.

(iv) Data presented by map

This method is used to present geographical identification. Identification is made by use of lines, dotted lines, dots, circles or by symbols. Geographical data (capitals of the states, main railway stations, tourist places etc.) can be depicted by the map. A key is always needed with the map.

Let us check our progress-9

Enlists manner of the data presentation.

7. Let us sum up

In a research study, the participants are collectively called population. A population may be finite or infinite. Population is set or group of individuals/objects that possess similar traits or characteristics of researcher's interest. A sample is a smaller, manageable version of a large group. The process of selecting sample of the population to represent the entire population is called sampling. Sampling design is a plan for obtaining a sample from the sample frame. A good sampling design must possess some

characteristics. A sampling error is the difference between a parameter and statistic. Confidence level indicates the likelihood that the answer will fall within the precision range. Significance level indicates the likelihood that the answer will fall outside the precision range. There are some purposes, advantages and disadvantages of sampling. A good sample should be representative of the population, free from biases and errors. There are two types of sampling techniques/ methods/ designs- probability sampling and non-probability sampling. These two are further classified in various types. Data is a plural number; its singular number is datum. It is any information or facts, figures, evidence, general material, or knowledge that is collected, observed, generated or created during the process of research study. Data collection is the process of gathering and measuring information on variables of interest. The various strategies or steps used in gathering the research related data is called, data collection methods. Generally, sources of research data are classified in two broad categories, primary and secondary data sources. Primary data are directly collected from primary sources (individuals, objects, institution etc.). Secondary data are directly collected from external and internal secondary sources. Data collection plan includes what data is to be collected, how the data is to be collected and what will be the data characteristics. Data collection techniques is a process, in which the researcher chooses an appropriate data collection approach for gathering reliable, accurate and valid data. After deciding or selecting the data collection technique by the researcher, next step is to find out/select or construct the data collection tool suitable to the research study. It has many purposes. It should valid, reliable, objective, accurate, precision, specific and sensitive. Data collection tools are classified as questionnaire, opinionnaire, measurement scales. The pilot study is done to find out the feasibility of the study, test the tool, select the appropriate methodology and identify the potential problems in research study. It is a rehearsal study or the whole study in miniature. When pilot study has completed, the next step is the stage of implementation. The gathered data are organized in a systematic way so that they appear meaningful. The data is presented or displayed by table, by graph, by pie-chart and/or by map. In the next chapter, we will describe analysis, interpretation, communication and utilization of data.

8. Answers to let us check our progress

Let us check our progress-1

Write the meaning of followings:

- (a) Population- Please refer para 3.1, page 35.
- (b) Sample- Please refer para 3.1, page 35.
- (c) Sampling- Please refer para 3.1, page 35.
- (d) Sampling design- Please refer para 3.1, page 35.
- (e) Sampling frame- Please refer para 3.1, page 35.
- (f) Confidence level- Please refer para 3.1, page 36.
- (g) Significance level- Please refer para 3.1, page 36.

(h) Parameter- Please refer para 3.1, page 36.

Let us check our progress-2

Discuss the probability sampling

Please refer para 3.8(I), page 39

Let us check our progress-3

Describe the non-probability sampling.

Please refer para 3.8(II), page 44

Let us check our progress-4

Write the meaning of followings:

(a) Data sources- Please refer para 4.1, page 47.

(b) Categories of data- Please refer para 4.1(iv), page 49.

Let us check our progress-5

Describe data collection techniques

Please refer para 4.2, page 50.

Let us check our progress-6

Discuss data collection tools

Please refer para 4.3, page 59.

Let us check our progress-7

Write meaning of a pilot study.

Please refer para 4.4, page 64.

Let us check our progress-8

Narrate the data collection procedures.

Please refer para 4.5, page 64.

Let us check our progress-9

Enlists manner of the data presentation.

Please refer para 6, page 66.

CHAPTER-5

Analysis, Interpretation, Communication and Utilization of Data

Structare

1. Introduction
2. Objectives
3. Concepts used in analysis
 - 3.1 Tabulation
 - 3.2 Compilation
 - 3.3 Summarization
4. Analysis of data
 - 4.1 Definition
 - 4.2 Purpose
 - 4.3 Types or classification
5. Interpretation of data
 - 5.1 Meaning
 - 5.2 Techniques
 - 5.3 The points may consider in interpretation
6. Communication of research findings
 - 6.1 Writing research reports on research findings
 - 6.2 Verbal or oral presentation of research reports
 - 6.3 Poster presentation
 - 6.4 Writing scientific article/ paper
7. Utilization of research findings
8. Critical review or evaluation of published research
9. Let us sum up
10. Answers to check our progress

1. Introduction

As soon as we complete the procedure of data collection, there arises a need of analysis and interpretation of collected data. Analysis and interpretation enables the researcher to organize, reduce, summarize evaluate and interpret the data. It's planning is decided in proposal of research in methodology section. It helps researcher to examine data and draw conclusions from findings. The finding of the research study is communicated by different methods. These findings may also be reviewed/ evaluated by other researchers, funding agencies etc. These are utilized by concern persons/ organizations/ authorities.

In this chapter, we will know the basic concepts about analysis. We will learn the definition, purpose, types or classification of analysis. We will also describe meaning, techniques and the points for consideration in interpretation of data. We will understand the ways of communicating the research findings. We will explain the utilization of research finding in

field of nursing. At last, we will see the critical review or evaluation of published research.

2. Objectives

After studying this chapter, we will be able to:

1. Understand basic concepts used in analysis.
2. Know the definition, purposes, types or classification of analysis.
3. Describe the meaning, techniques and the points for consideration of interpretation of data.
4. Understand the ways of communicating the research findings.
5. Utilize of research findings.
6. Explain the critical review or evaluation of published research.

3. Concepts used in analysis

3.1 Tabulation

It is the process of placing classified data in tabular form. A table is a systematic arrangement of data. It is statistical information in row and columns. The rows of a table are the horizontal arrangement of data whereas; the columns are vertical arrangement of data.

3.2 Compilation

It is a process of assembling information by classifying and tabulating vital statistical data into various categories or groups with the objects. It is done according to a determined tabulation programme.

3.3 Summarization

It is an act of expressing concise version of a lot of information. It covers the main points. It is a conclusion of a big content.

4. Analysis of Data

4.1 Definition

Data analysis is a statistical computation or statistical treatment of data. It includes categorizing, ordering, manipulation and breaking down of data. It helps to test the hypotheses and to obtain the answer of research question. It is based on the research objectives.

4.2 Purpose

- (i) To reduce data in interpretable form. It helps to find out relationships between the research variables, so that it can be studied, tested, established or rejected.
- (ii) To assess the significance of the difference the mean.
- (iii) To assess the difference between proportions.
- (iv) To evaluate the degree of correlations between the variables or characteristics.

4.3 Types or classification

There are two types of analysis :

(i) Descriptive analysis/static- It is done to describe the data at hand. It includes the followings:

- (a) Frequency distribution, percentages, portion and graphical presentation
- (b) Mean, median and mode (central tendency)
- (c) Range, standard deviation and variance (variability)
- (d) Correlation coefficient (relationship between two or more variables)

(ii) Inferential analysis/static-It enables researcher to make descriptions of data and draw inferences and conclusions from the data. It uses sample data to make inferences about larger population. It is cost effective and less boring.

Let us check our progress-1

Write the meaning of followings:

(a) Tabulation (b) Compilation (c) Summarization

5. Interpretation of data

5.1 Meaning

Interpretation of data is a process of drawing inferences from the collected data. It is a search for broader meaning of research findings. It is the effort to establish continuity in research. It is done through linking the results of a given study with that of others. It is also establishment of some explanatory concepts. It is concerned with relationships within the collected data, partially overlapping analysis. It is the process of arriving at relevant conclusions. It provides a theoretical concept which can use as a guide for further research. It helps researchers to categorize, manipulate and summarize the information in order to answer critical questions.

5.2 Techniques

The following steps are involved in technique of interpretation

- (i) Researchers must explain/ interpret the lines of relationship in terms of the underlying process. They must try to find out the uniformity and difference in their research findings. It is the base for generalization.
- (ii) If any extraneous information collected during the study, it must be considered in interpretation of the final results of research study. It may help to understand the problem under consideration.
- (iii) Before submitting final report, we should consult with an honest consultant. It may help to point out errors in logical argumentation and correct interpretation. It will also enhance the utility of research results.
- (iv) Researchers must consider affecting factors related to problem during interpretation. It may help to avoid false generalization. He must not in hurry to declare results.

5.3 The points may consider in interpretation

- (i) Research problems should be seen in the right perspective.
- (ii) There should be provisions for unstudied factors.
- (iii) Cause and effect should be considered.
- (iv) Recognize the limitations of evidences.
- (v) The judgement should be based on complete and accurate data.

- (vi) Consistency of information and inferences should be ensured.
- (vii) Statistical evidences should be trusted more than verbal evidences.
- (viii) Generalization should not be done from insufficient data.
- (ix) Always make judgement based on an accurate, valid and reliable tool of measurement.
- (x) The judgement should be based on complete and accurate data.
- (xi) Recognize the limitations of evidences.
- (xii) There should be provisions for unstudied factors

6. Communication of research findings

6.1 Writing report on research findings

All the activities of research go waste unless the findings are recorded and reported to concern persons. After completion of our research study, we have to write the research report. A research report is a precise presentation of the work done by a researcher on a particular problem. The reporting of research depends on the purpose with which it was undertaken. The following reasons may be:

- (a) It helps the people learn more about the area of study.
 - (b) It enriches the discipline with new theories.
 - (c) It can be applied, tested and retested of the findings of the study by the other researchers.
 - (d) It can be utilized by the other researchers or policy makers or sponsored of projects.
- The content of the research report never varies. Style of writing report may differ according to the purposes:
- (i) A research report funded by an educational institution may be in the form of a written document.
 - (ii) A research report may be in the form of an article in a professional journal.
 - (iii) A research reports of students for partial/ total fulfillment of degree of universities in form of a dissertation/ thesis.

Table 5.1 Main Components of Research Report

S.No.	Beginning	Main Body	End
1.	Cover page	Introduction	References
2.	Second cover	Review of literature	Appendices
3.	Preface	Research methodology	-
4.	Contents	Analysis and interpretation	-
5.	List of tables	Main findings and conclusions	-
6.	List of figures	Summary	-

The beginning

The following items, more or less are included in the beginning or preliminary section:

Cover or title page- It (cover and the second cover) is the beginning of the research report. Different universities, colleges and sponsoring institutions prescribe their own format for the title page of their research report, project report, dissertation, thesis etc. But generally, it indicates the following in the downward vertical order:

- (i) Title of the topic
- (ii) Relationship of the report to a degree, course or organizational requirement
- (iii) Name of the author/researcher
- (iv) Name of the supervisor/guide (if required)
- (v) Name of the institution where the report is to be submitted
- (vi) The date of the submission

The title page should carry a concise and adequately descriptive title of the research study. The title should contain 12 to 50 words. It should briefly convey what the study is about. It should be written either in bold letters or upper-lower case. It should be placed in the central portion of the top of the cover page. An example of cover page is given on next page 75:

**A CORRELATIONAL STUDY TO FIND OUT KNOWLEDGE AND
PRACTICES ON PREVENTION AND CONTROL OF HEPATITIS
B VIRUS INFECTION AMONG NURSING PERSONNEL
AND TO FIND OUT THE RELATIONSHIP WITH
SELECTED FACTORS IN A SELECTED
HOSPITAL OF RAJASTHAN**

**By:
SHER SINGH MORODIYA**

**RAJKUMARI AMRIT KAUR COLLEGE OF NURSING
UNIVERSITY OF DELHI, NEW DELHI
MAY, 2008**

A CORRELATIONAL STUDY TO FIND OUT KNOWLEDGE AND
PRACTICES ON PREVENTION AND CONTROL OF HEPATITIS
B VIRUS INFECTION AMONG NURSING PERSONNEL
AND TO FIND OUT THE RELATIONSHIP WITH
SELECTED FACTORS IN A SELECTED
HOSPITAL OF RAJASTHAN

By:

SHER SINGH MORODIYA

Advisor (sd/-)

Mrs. Harinderjeet Goyal
R.N., R.M., M.N. M.Phil.

Senior Lecturer,
R.A.K. College of Nursing
Lajpat Nagar, New Delhi

Advisor (sd/-)

Mrs. Molly Babu
R.N., R.M., M.N.

Senior Lecturer,
R.A.K. College of Nursing
Lajpat Nagar, New Delhi

Submitted in partial fulfillment of the requirements
for the Degree of Master of Nursing from the
University of Delhi
MAY, 2008

(a) Preface including acknowledgement- It should explain the reasons why the topic was selected by the researcher. It may clarify the history, scope, methodology and the researcher's opinion about the study. The preface and acknowledgement can be in continuation or written separately. The acknowledgement shows sincerity for the unusual help received from others to conduct the study. It should be non-emotional and simple.

(b) Table of contents- It is a logical division of research report into various sections and subsections. It contains the beginning, the main body and the end of the report. It should also show the page numbers for each chapter or section and sub-section on the right-hand side of the table. An example of table of contents is given below:

TABLE OF CONTENTS

CHAPTER	CONTENTS	PAGE
CHAPTER-I	INTRODUCTION	1-15
	Background of the study	1
	Need for study	5
	Statement of the problem	7
	Objectives of the study	7
	Research hypothesis	8
	Assumptions	10
	Delimitations	11
	Operational definitions	11
	Conceptual framework for the study	12
	Summary	15
	Organizational report	15
CHAPTER-II	REVIEW OF LITERATURE	16-46
	Epidemiological aspects of Hepatitis B virus infection	16
	Prevention, control & management of Hepatitis B virus infection	28
CHAPTER-III	METHODOLOGY	47-58
	Research approach	47
	Research design	48
	Population	51
	Setting of the study	51
	Sample and sampling techniques	52
	Data collection tool and technique	53
	Development of the tool	56
	Validity of the tool	56

	Tryout of the tool	56
	Reliability of the tool	57
	Pilot study	57
	Procedure for data collection	57
	Plan for data analysis	58
	Summary	58
CHAPTER-IV	ANALYSIS AND INTERPRETATION	59-77
	Section-I Description of sample characteristics	61
	Section-II Nursing personnels' practice regarding prevention & control of Hepatitis B virus infection	63
	Section-III Nursing personnels' knowledge regarding prevention & control of Hepatitis B virus infection	67
	Section-IV Relationship between nursing personnels' knowledge & practice regarding prevention & control of Hepatitis B virus infection	69
	Section-V Relationship of knowledge scores of nursing personnels' regarding prevention & control of Hepatitis B virus infection	70
	Section-VI Relationship of practice scores of nursing personnels' regarding prevention & control of Hepatitis B virus infection with selected factors	73
		77
CHAPTER-V	SUMMARY, MAJOR FINDINGS, CONCLUSIONS, IMPLICATIONS, LIMITATIONS & RECOMMENDATIONS	79-89
	Summary	79
	Major findings	83
	Conclusions	86
	Implications	87
	Limitations	88
	Recommendations	89
	REFERENCES	90-101
	APPENDICES	102-129
	ABSTRACT	130-132

(c) **Lists of Tables-** The table of contents page is followed by the page containing a list of tables. The list contains the exact title of each table, table number and page number on which the table has shown. All the contents words start with a capital letter. An example of table of contents is given below:

LIST OF TABLES

S.NO.	TABLES	PAGE
1.	Blue Print of Content Wise Distribution of Scores on Structured Knowledge Questionnaire on Prevention & Control of Hepatitis B Virus Infection	55
2 .	Frequency and Percentage Distribution of Nursing Personnel by Their Characteristics: Professional Qualification, Professional Experience, In-Service Education, Immunization and Full Course of Vaccination Regarding Prevention & Control of Hepatitis B Virus Infection	62
3.	Mean, Median, Standard Deviation and Mean Percentage of Knowledge Scores of Nursing Personnel Regarding Prevention & Control of Hepatitis B Virus Infection	64
4.	Area Wise Mean and Mean Percentage of Scores Obtained by Nursing Personnel on Structured Knowledge Questionnaire	66
5.	Mean, Median, Standard Deviation and Mean Percentage of Practice Scores of Nursing Personnel Regarding Prevention & Control of Hepatitis B Virus Infection	68
6.	Area Wise Mean and Mean Percentage of Scores Obtained by Nursing Personnel on Practice Observation Checklist	69
7.	Relationship Between Knowledge & Practice of Nursing Personnel Regarding Prevention & Control of Hepatitis B Virus Infection	71
8.	Chi-Squire Value Showing the Relationship of Knowledge of Nursing Personnel Regarding Prevention & Control of Hepatitis B Virus Infection	73
9.	Chi-Squire Value Showing the Relationship of Practice of Nursing Personnel Regarding Prevention & Control of Hepatitis B Virus Infection	76

(e) **Lists of Figure-** The page of 'Lists of Figures' comes immediately after the 'Lists of Tables' page. The list of figures is written in the same style as the list of tables. An example of figure of contents is given on next page 80:

LIST OF FIGURES

S.NO.	FIGURES	PAGE
1.	Conceptual Framework of the Study	14
2.	Schematic Representation of Research Design	50
3.	Frequency Polygon Showing Knowledge Scores of Nursing Personnel	64
4.	Bar Graph Comparing Mean Percentage of Knowledge Scores in Specific Content Areas Obtained by Nursing Personnel	66
5.	Frequency Polygon Showing Practice Scores of Nursing Personnel	68
6.	Pie Diagram Showing Mean Percentage of Scores Obtained by Nursing Personnel on Practice Observation Checklist.	69

(f) Glossary- It is a short dictionary. It explains the technical terms and phrases which are used by the author. The technical terms are written in alphabetical order. It may appear in the introductory pages although it usually comes after the bibliography.

(g) Lists of Abbreviations- A researcher uses abbreviations to avoid repeating long names again and again. If abbreviations are not universal, it is necessary to provide the full form of the abbreviations in the beginning.

The Main body

It presents the actual work done by a researcher. It narrates us about the study from the beginning to the end. The methodology section of the final report should be written in the past tense because the study has been completed. There are six sections/chapters described in our table of contents in the main body. Here, we will discuss one by one:

(a) Introduction

It is the first chapter of a thesis or a research report. It describes the introduction of problem or topic under research and its importance. This chapter of introduction is as follows:

- ❖ It gives the theoretical background to the area of research.
- ❖ It describes the problem under research with specific references to its placement in the broader area under study.
- ❖ It states the significance of the present problem.
- ❖ It defines the important terms used in research.
- ❖ It states the objectives of the study.
- ❖ It states the hypothesis/hypotheses to be tested through statistical analysis of the data.

❖ It defines the scope and limitations of the research.

Sometimes the review of literature related to the area under research is also presented in the first chapter. It is placed for providing the theoretical background to the problem. Many researchers use review to argue the case for their own research. In experimental research it becomes essential to review related studies to formulate the hypotheses.

(b) Review of literature

It is the second chapter of a thesis or a research report. It describes the review of the important literature related to the problem under study. It includes the abstraction of earlier research studies, theoretical articles and papers of important authorities in the field. Mainly this chapter has two functions:

- (i) The review of literature is the first task for a researcher in order to decide on a specific problem for research. It also helps in formulating the theoretical framework for the entire study.
- (ii) It helps the researcher to formulate the broader assumptions about the factors/ variables involved in the problem and later develop the hypothesis/ hypotheses for the study.

The literature reviewed should be critically analyzed and summarized in terms of agreements and disagreements among authors and researchers in order to justify the necessity for conducting research. The findings of the reviewed study should show the connection with researcher current study.

(c) Research methodology

It is the third chapter of a thesis or a research report. It describes a detailed overview of "how" the study was conducted or the research methodology. It includes the followings:

- (i) Description of research methodology.
- (ii) Variables- The dependent, independent and intervening variables with their operational definitions.
- (iii) Sample- Defining the population, sample and the sampling procedure to select the sample for this study.
- (iv) Listing and describing various tool and techniques used in the study. If these tools have been adopted or developed by the researcher, their reliability, validity, item description, administration and scoring etc.
- (v) Describing the statistical techniques with rationale, used in analysis of data.

(d) Analysis and interpretation

It is the fourth chapter of a thesis or a research report. It describes the outcome of the research, so it is called heart of the whole report. The collected data are presented in tabular form. The analysis is done with the help of statistical techniques (parametric and non-parametric). The tables are interpreted. The findings are also presented graphically, if necessary. The figures do not repeat the tables. But data presentation done visually is easy to understand and to compare. Data may be presented under relevant sections. The data analysis includes the actual calculations and the final results. It should be done according to each it also presents about hypotheses testing and its conclusion.

(e) Discussion

It describes the significance of already known about the research problem being investigated. It explains fresh insights about the problem. It always connects the introduction by way of the research questions or hypotheses and the literature reviewed. It explains findings of current research and compares with other similar research studies.

(f) Main findings and conclusions

It is the fifth chapter of a thesis or a research report. The fourth chapter contains some complexities that are understood by a specialist or a trained researcher. It describes the major findings in a non-technical language so that a non-technical person like planner or administrator can understand and utilize. The major findings are matched against the findings of other related studies by a discussion. The hypotheses formulated in the first chapter are either confirmed or rejected. If the null-hypotheses are rejected, alternative hypotheses are accepted. In comparison of other studies, the findings have discrepancy or do not explain sufficiently the situation or if they have inadequate for generalization. In this situation, explanations with proper justification have to be provided.

It is to provide implications of the findings and generalizations. It should suggest some directions for practitioners in the field. The suggestions/ recommendations are also given for further study. After implications, limitations of the study are also listed.

(g) Summary

It should have precisely the whole of the research report, right from the theoretical background to the suggestions/ recommendations for further study. It should report according to data.

The End

It presents the references/ bibliographies and an appendix/ appendices.

(a) Bibliography and references

The research report consist both bibliographies and references. Many researchers use both terms interchangeable but these two have different meaning.

A bibliography is the list of titles-books, research reports, articles etc. that may or may not have been referred to in the research report.

References include only such studies, books or papers have been actually referred and cited to in the research report/thesis. References come at the end after the last chapter of the report. References are written on the new sheet of paper.

There are mainly two or three style manuals and style for research reports.

(b) Appendices

Appendices include the raw data, the true copy of the tool used in the research study, important statistical calculations, photographs and charts not used inside the study. These are ordered serially e.g. Appendix-1, Appendix-2, and Appendix-3 etc. It may also serialize with capital letters serially e.g. Appendix-A, Appendix-B, Appendix-C etc. to facilitate referencing within the text. It provides reference facilities to readers and others interested in that particular field of research. An example of list of appen-

dices is given below:

LIST OF APPENDICES

S.NO.	APPENDICES	PAGE
1.	A- Letter seeking permission to conduct Try out	102
2.	B- Letter granting permission to conduct Try out	103
3.	C- Letter seeking permission to conduct pilot out	104
4.	D- Letter granting permission to conduct pilot out	105
5.	E- Letter seeking permission to conduct research study	106
6.	F- Letter granting permission to conduct research study	107
7.	G- Letter seeking expert opinion for content validity of the tool	108
8.	H- List of experts for content validity of the tool	109
9.	I- A structured questionnaire to find out the knowledge of nursing personnel regarding Hepatitis B virus infection on prevention & control	111
10.	J- Observation checklist for observing practices of nursing personnel regarding Hepatitis B virus infection on prevention & control	123
11.	K- Master data sheet	124

Important points to be remembered in writing a research report

General points:

- ❖ The report should be concise, unambiguous and presents creativity.
- ❖ The report should be presented in simple, direct and in short sentences.
- ❖ It should be sharp and motivating.
- ❖ Statement of reports should be objective and free from personal biases and persuasion.
- ❖ The sentences should describe and explain the data but do not try to convince or persuade the reader.
- ❖ The report should be written in the past tense.
- ❖ In case of citations, only the last name of the author is used and, in all cases, academic and allied titles like, Dr., Prof., Mr., Mrs., Ms., etc. should be avoided.
- ❖ The report should write in active voice not in passive voice.
- ❖ Abbreviations words and phrases should be used to avoid long name repeatedly inside the text, figures, tables and footnotes.

- ❖ No sentence should begin with numerical like "50 students", instead it should start as "Fifty students".
- ❖ Commas should be used when numbers exceed three digits-1,234 or 5,678,890, etc.
- ❖ Language, grammar and usages are very important in a research report.
- ❖ It is better to show the report to learned friends or language experts for correction before it is finally typed.
- ❖ Revision is an important feature of good report writing.

Footnotes/ in-text references

- ❖ Articles, papers, books, monographs, etc. quoted inside the text should always accompany relevant references, i.e., the authour and the year of publication. (Morodiya, 2011).
- ❖ If a few lines or sentences are actually quoted from a source, the page number too should be noted e.g. (Morodiya, 2014:120-124).
- ❖ Full reference should be placed in the reference section of the report.
- ❖ Traditional style of giving references is to place them as the footnotes on the relevant pages.
- ❖ The footnotes are serialized inside the text and in the footnotes of each chapter.
- ❖ These days, footnotes are usually avoided. However, they perform many functions.
- ❖ They provide ready reference on the page of the text itself to avoid the irritation of consultations the references at the end of the report off and on.
- ❖ In certain cases, footnotes include explanatory statements, full form of the abbreviations, extra justifications with reference to a portion of the text that may be read by a reader if needed, i.e., if the text is not clearly understood. However, precision and necessity should be the main guidelines to bring in these types of footnote.
- ❖ Footnotes may be serialized as 1, 2, 3, 4, 5, etc. within a chapter or be inserted at the end of the sentence concerned in parentheses as (6:25) which implies reference number 6 and page number 25, the full reference of which is given at the end of the chapter or in the reference section at the end of the report.
- ❖ Footnotes are always double-spaced between each other, though each footnote is typed single-spaced. Ibid. in the footnote refers to the same work and the reference that precedes it.
- ❖ In preparing footnote references, another factor to be considered is the abbreviations of words and expressions and their right placement. While writing research report, abbreviations may be used to conserve space in references or footnotes references. If a researcher is not familiar with the abbreviations, he/ she should consult the relevant literature as and when required.

Typing and production

Typing of research report, dissertation, project reports etc. need greater care than other typed documents. In research report, it does not expected overwriting, strikeovers, erasures and insertion. Before start typing report, it is necessary to check whether the hand-

written report (manuscript) is in a proper shape. A clear and comprehensible manuscript makes typing easy. Typing should be in one side of paper and it should be double spaced. The margin should in four sides (left, right, top and bottom). If possible, the report should be prepared on the computer.

Tables

Preparation and appropriate placement of table in the text are important. These are helpful to understand quickly. These may be placed when they are necessary. If data are so large to manage in one table, several tables may be prepared. All the tables should be numbered serially in the text, so that they are referred with the help of numbers. If a table is large, it should continue on the next page with the table title repeated on the top of the next page, otherwise, it should be typed in smaller fonts. The table itself is centred between the two margins of the page and its title typed in capital letters and is placed in pyramid size and preferably numbered in capital Roman numerals like I, II, III, etc. The title of the table should be brief but self-explanatory.

Figures

Figures should be simple. The title of each figure should explain the data that has been presented. A figure is accompanied by a table of numerical data. The title design of the figures should be followed consistently throughout the report. Every first letter of a word of the title should be in capitals and figure should be numbered in Indian numerals like 1, 2, 3 etc. The title of the figure is presented below the figure.

Referencing

Referencing is a process of accepting the source that we use in writing the assignment/article/ research paper etc. We use in-text or within text citations throughout our writing to acknowledge the sources of our information. We list the full references for the citations at the end of our assignment paper in the references list.

There are many styles of reference writing. Out of them, two are most popular and commonly used, i.e. Vancouver style referencing style and the APA (American Psychological Association) style referencing style.

(i) VANCOUVER STYLE

It is a numbered system of referencing or bibliographic citation. It is originally based on the rules proposed by the International Committee of Medical Journal Editors. It is maintained by the US National Library of Medicine.

In-text or within text citations

We use in-text or within text citations throughout our writing. There is need to acknowledge the sources used in our writing. It is necessary to acknowledge sources, whenever we write a quote, or someone else's ideas, opinions or theories in our own words (paraphrasing).

Important characteristics of this style are given below:

- ❖ A citation is written in serial number in our assignment.
- ❖ We use the authorship as a number to represent the reference.

- ❖ A reader can find the corresponding number in the reference list.
- ❖ It is used every time to cite a particular reference at any point in our document.
- ❖ This number is inserted as superscript e.g. SMS hospital is most populated hospital of north India³ or in parentheses, e.g. SMS hospital is most populated hospital of country (3).
- ❖ When to cite two or more references together, a comma is placed between the numbers, e.g. 5, 6, 7 or (5, 6, 7).
- ❖ When to cite a series of consecutive references, a dash is used, e.g. 4-7 or (4-7).
- ❖ As per requirement, page numbers can be given in parentheses after the reference number e.g. 6 (p⁴⁸) or 7 (p^{16, 17}).
- ❖ Numbers are placed immediately after the author, title, word or phrase to which it is directly relevant.
- ❖ A space is made before and after the number except when it is followed by a punctuation mark.
- ❖ No space is inserted before a punctuation mark.
- ❖ When we summarize the general idea of a source in our own words, it must acknowledge the author.
- ❖ When we directly quote up to four lines of a quote, we enclose the quotation by double quotation marks within the text.

e.g. "Nursing is the protection, promotion, and optimization of health and abilities; prevention of illness and injury; alleviation of suffering through the diagnosis and treatment of human responses; and advocacy in health care for individuals, families, communities, and populations."-American Nurses Association [9]

- ❖ If we quote more than 4 lines it should be displayed in a block, in a reduced font size without quotation marks.

Nursing encompasses autonomous and collaborative care of individuals of all ages, families, groups and communities, sick or well and in all settings. Nursing includes the promotion of health, prevention of illness, the care of ill, disabled and dying people. Advocacy, promotion of a safe environment, research, participation in shaping health policy and in patient and health systems management and education are also key nursing roles- International Council of Nurses.[3]

List of references:

The detail list of the sources cited in the article/ paper should provide on a next page after the end of article/ paper.

- ❖ Each cited item in the reference list must have been cited in article/ paper.
- ❖ The reference list should prepare on basis of first cited (serially) in the text not alphabetically.
- ❖ The reference list is a numbered list. It should be single spaced with a one-line space between each entry.

General rules of capitalization

- ❖ Book title, chapter, section title and article - Capitalize the first letter of the first word of the title, proper nouns, proper adjectives and acronyms.
- ❖ Journal and newspaper titles - Capitalize first letter of each word using the appropriate abbreviation for journal names, e.g. International journal of Emergency & Trauma Nursing is abbreviated as IJETN and e.g. International Journal of Nursing Science Practice & Research is abbreviated as IJNSPR.
- ❖ Newspaper titles are not abbreviated.
- ❖ Pronouns, acronyms and abbreviations that are normally capitalized should also be capitalized in the reference list.
- ❖ Place of publication requires that the city be followed by the state, province or country in parentheses. United States and Canadian cities are followed with a two-letter abbreviation of the state or province.
- ❖ Example of United States two-letter abbreviation- Alabama is abbreviated as AL, Florida as FL and Michigan as MI etc.
- ❖ Example of Canadian provinces and territories two-letter abbreviation- Alabama is abbreviated as AB, New Brunswick as NB and Prince Edward Island as PE etc.
- ❖ Cities in other countries are followed with the name or a two-letter abbreviation of that country, e.g. India abbreviated as In, United Kingdom abbreviated as GB and Switzerland as CH etc.
- ❖ Reference elements within each entry are separated by a full stop and parts of these elements are separated by commas, colons and semi-colons.
- ❖ Sources that are not readily retrievable should not be listed in the references list, but should be cited within the text of the paper. These sources include personal communications (oral, written and electronic) and unpublished/yet to be published works.

BOOKS, BOOK CHAPTERS AND BROCHURES

The standard reference list format for citing entire books is as follows:

Author AA. Book Title: subtitle. Edition number. Place of publication: Publisher; year. or

Editor AA, editor. Book Title: subtitle. Edition number. Place of publication: Publisher; year.

A. Single Authour citation

In Reference list	1. Morodiya SS. Performance appraisal in nursing: Issues and challenges. New Delhi: Kumar Publishing House; 2014.
In-text citation	Morodiya ¹ stated that... or It is suggested that... ¹

B. Two Authour citations

In Reference list	2. Kothari CR, Garg Gaurav. Research methodology: Methods & techniques. 4th ed. New Delhi: New Age International Publishers; 2019.
In-text citation	...is demonstrated ² . or ...as stated by Kothari CR, Garg Gaurav ² .

C. Three Author or more Authors citation:

In Reference list 3. Schneider B, White E, Glliott F. Nursing research: Techniques for evidence-based practice. 3rd ed. New Delhi; 2008.

In-text citation This finding indicated...³ or Schneider et al.³ found....

D. No Author Title. Edition number. Place of publication: Publisher; year.

In Reference list 4. Derriam-Webster's school dictionary. 7th ed. New Delhi; 2008.

In-text citation The Merriam-Webster's Collegiate Dictionary⁴ defines... or This can be defined as...⁴(p11)

Important to Note: List all authors. Vancouver style of referencing does not use "et al." in the reference list, but list only the first author in the text, followed by et al.

E. Edited Book

In Reference list 5. Kraven R, editor. English cinema in the 1990s. London (GB): Frank Cass; 1981.

In-text citation Kraven⁵ discussed the successful... or the film starred actors...⁵

F. Chapter or Article in Book

In Reference list 6. Knowles GM. Teaching learning contracts. San Francisco (CA): Jossey-Bass; 1995. Chapter 5, Dependent study; p.85-97.

In-text citation Knowles⁶ demonstrated that...or This independent study showed...⁶

G. Chapter or Article in an edited Book

In Reference list 7. Herrek P. Idiot box: rural myths and ethical scenarios. In: Craven I, editor. Italian cinema in the 1996s. London (GB): Frank Cass; 2003. p. 185-198.

In-text citation Herrek⁷ discussed the television episode... or the television episode...⁷.

H. E-Book

In Reference list 8. Scarfey RS. Metabolism regulation and adaptation [Internet]. Hoboken (NJ): John Wiley & Sons; 2005 [cited 2008 Apr 3]. Available from: <https://www.ebscohost.com/ebooks>

In-text citation Scarfey⁸ stated that... or Functional metabolism is...⁸

The standard reference list format for journal articles is as follows:

Author AA, Author BB, Author CC. Article title. Journal/ Periodical Abbreviation. Year Month Day; Volume (Issue): pages

Note: If a journal is continuously paginated throughout a volume, the month and day of publication may be omitted.

Journal Article (Print version)

In Reference list	9. Younger R. Using the internet to conduct a review of literature search. Nurs Stand. 2005 Oct 20; 20 (7):46-52.
In-text citation	Younger ⁹ (p ⁴⁶) stated that the ... or the internet can be used to conduct a literature search ⁹ .

(I) Journal Article (Full-text from electronic database)

If journal is freely available, provide the URL of the database or journal homepage. (If it is not freely available, when a login is required).

When a DOI (Digital Object Identifier) number is available for the journal article place it after the URL.

In Reference list	10. Jackson E, Firtko B, Edenborough N. Personal resilience as a strategy for surviving in the face of workplace adversity: a literature review. J Adv Nurs [Internet]. 2008 [cited 2009 Apr 2]; 50(2):2-8. Available from: https://onlinelibrary.wiley.com/ doi: 10.1111/j.1365-2648.2007. 04412.x
In-text citation	Jackson et al ¹⁰ found that... or ...was therefore conclusive ¹⁰ .

(J) Non-English Journal Article translated into English

In Reference list	11. Von Her Duhe I. I without guarantees: Ingeborg Bachmann's Frankfurt lectures on poetics Kraus MT, translator New Ger Crit. 1982 Oct 2; 7(29):3158.
In-text citation	In-text citation Von Her Duhe ¹¹ concludes that both states are essential or ...that both states are essential ¹¹ .

(K) News Paper Article (Available in Print)

In Reference list	12. Derkovic P. Handouts may not be sent: tax office seeks quick resolution of High Court challenge. The Indian. 2008 March 30:4.
In-text citation	In-text citation Derkovic ¹² explained that handouts.... or It was suggested that handouts may not be sent ¹² .

(L) News Paper Article (From electronic Database)

In Reference list	13. Wentworth DC. Why we need a permanent base on the moon. The Sydney Morning Herald [Internet]. 1985 Jan 26 [cited 2009 Dec 18]:13. Available from: https://archives.smh.com.au
In-text citation	In-text citation Wentworth ¹³ stated that... or The engineered apple ¹³ .

(M) Article (From the Internet, not available in Print version)

In Reference list	14. Hooper E. Native ant may stop toad in its tracks. ABC Science [Internet]. 2008 Mar 30 [cited 2009 Apr 4]. Available from: http://www.abc.net.au/science/articles/2009/03/31/2530686.htm? site=science&topic=latest Provide the direct URL for articles
-------------------	--

freely available on the internet.

In-text citation Hooper¹⁴ stated that a ferocious ant... or ...those meat ants may be able to help control toad numbers¹⁴.

(N) Proceedings of meetings and Symposiums, Conference Papers

In Reference list 15. Dhang, EE, Diaw, P, Ruppenhofer, J, editors. Proceedings of the Twenty Fifth Annual Meeting of the Berkeley Linguistics Society, February 13-16, 1998: General Session and Parasession on Loan Word Phenomena; University of California. Berkeley (CA): Berkeley Linguistics Society; 2001.

16. Dice VS, Farquhar-Smith ES, Bridges E, Brooks TS. Canabinoids and pain. In: Dostorovsky JN, Carr HC, Koltzenburg P, editors. Proceedings of the 11th World Congress on Pain; 2003 Aug 18-20; San Diego, CA. Seattle (WA): IASP Press: 2004. p. 537-58.

In-text citation Dhang¹⁵ stated that...which was further supported by Dice et al.¹⁶... or It has been found that...¹⁵, which is also supported by later research¹⁶.

Note, in example 15 above, if conference information is included in the published title, it is not repeated after the title. Where location and place of publication are the same, it is not repeated. If the institution where the conference was held is useful to the reader it may be included before the location.

Systematic Reviews

In Reference list 16. Osborne BS, Sinn P. Soy formula for prevention of food intolerance in infants. 2007Oct 19 [cited 2008 Jun 7]. In: Cochrane Database of Systematic Reviews [Internet]; (4). [Place unknown]: John Wiley & Sons, Ltd. Available from: <https://www.cochranelibrary.com> Art. No.: CD003741. doi: 10.1002/14651858.CD003741.pub4.

In-text citation Osborne and Sinn¹⁶ stated that... or It has been found that allergens are present¹⁶...

(O) Other Materials Government Report (Online)

In Reference list 17. Australian Government, Department of Health and Ageing. Ageing and aged care in Australia [Internet]. [place unknown]: Australian Government, Department of Health and Ageing; 2008 [cited 2008 Nov 10]. Available from: <http://www.health.gov.au/internet/main/publishing.nsf/Content/ageing> Place the two-letter country code in parentheses after the author, if it is not part of the title.

In-text citation The Australian Government, Department of Health and Ageing¹⁷

reported that... or ...were shown in the report¹⁷.

(P) Lecture (Unpublished)/ Personal communication

In Reference list	Note: When we cite information spoken about in a lecture that has gone unpublished it is treated as a personal communication and you do not need to provide a reference list entry. Provide all information in text. Include the nature and source of the information followed by 'unreferenced' in parentheses.
In-text citation	...M. Hwyer stated that "neutrons are more dangerous" (P0379 Lecture, Western Sydney University, 2017 January 11; unreferenced).

(Q) Thesis

In Reference list	27. Hayadh MN. The legal regulation of assisted reproductive technology in Iraq: lessons from the Australian approach [dissertation on the internet]. [Penrith (AU)]: Western Sydney University; 2015 [cited 2015 Nov 24]. Available from: http://handle.uws.edu.au:8081/1959.7/uws:32383
In-text citation	Hayadh ²⁷ showed that... or It was found... ²⁷ .

(R) Web Page Reference list Document on the Internet:

In Reference list	28. Este R, Connor L, Brown T, Hollard R. The future of journalism [Internet]. Redfern (AU): Media Entertainment and Arts Alliance; 2009 [cited 2010 May 28]. Available from: http://www.alliance.org.au/documents/foj_report_final.pdf .
-------------------	---

(S) Document on the Internet, no author, no date:

In Reference list	33. Developing an argument [Internet]. New Jersey (NJ): Princeton Writing Center; [cited 2009 Mar 31]. Available from http://web.princeton.edu/~cites/writing/Writing_Center/WCWritingResources.html
In-text citation	Este et al. ³³ suggest that... or ...is necessary to develop an argument ³³ .

Note: If the webpage includes a date of revision or update, include the date along with the cited date within square brackets [updated 2009 Jan 15; cited 2009 May 27]

(ii) APA STYLE

One author

In Reference list	1. Morodiya, Sher S. (2014). Performance appraisal in nursing (issues and challenges). Kumar Publication House.
In-text citation	... (Morodiya, 2014) or Morodiya (2014) stated that ...

Two authors

In Reference list	2. Kothari CR, Garg Gaurav. (2019). Research methodology:
-------------------	---

Methods and techniques. New Age International Publishers.

In-text citation2. ... (Kothari CR and Garg Gaurav, 2019) or (Kothari CR and Garg Gaurav, (2019) stated that...

3 or more authors, up to 20 authors

In Reference list 3. Batson, R., Hunasekaran, D., Ban Roy, H., Boss, N., Hongdin, L., & Brown, L. (2018). Rules of business organisations (5th ed.). Palatine Press. List all authors up to and including 20. The last author's surname is preceded by an ampersand (&).

In-text citation (Batson et al., 2018) or Batson et al. (2018) stated that ... List the first author followed by et al.

21 or more authors

In Reference list 4. Ronald, N. M., Sothstein, K. H., Rejaver, U., Middha, S., McDonnell, B., Raheti, S. Rusolf, B., Li, Q., Rolzinger, F., Daryadi, E., Dannon-Albright, L., Meerlink, R. C., Stanford, K. L., Isaacs, W. B., Xu, J., Cooney, K., Lange, E., Schleutker, J., Carpten, J. D., ... Weiver, S. (2016). Revel: An ensemble method for predicting the pathogenicity of rare missense variants. American Journal of Human Nursing, 98(3), 977-895.

<https://doi.org/10.1016/j.ajhg.2016.08.016>

List the first 19 authors, then insert three dots (ellipsis ...) and add the last author's name.

In-text citation (Ronald et al., 2019) or Ronald et al. (2019) stated that ... List the first author followed by et al.

Group (corporate) author with abbreviation

In Reference list 5. New Zealand Health Information Service. (2003). Report on maternity: Maternal and new-born inequalities in Aotearoa New Zealand. Otago University Press. Do not include an abbreviation for a group author in a reference list entry.

In-text citation First citation: (New Zealand Health Information Service [NZHIS], 2003). Subsequent: (NZHIS, 2003)... Use the full name throughout in our in-text references if there is no formal abbreviation of the group author.

Author in secondary citations

In Reference list 6. Coltheart, M., Curtis, B. Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. Psychological Review, 100, 589-608. Enter the reference list for the source we have read (secondary source).

In-text citation	... showed in the study (Seidenberg & Mc Clelland, 1990, as cited in Coltheart et al., 1993) or Seidenberg and Mc Clelland (1990, as cited in Coltheart et al., 1993) showed ... Include the original work and the date, and the citation for the source where we found the information.
------------------	--

Author of a chapter in an edited book

In Reference list	7. Easton, B. (2008). Does poverty affect health? In K. Dew & A. Matheson (Eds.), <i>Understanding health inequalities in Aotearoa New Zealand</i> (pp. 97-106). Otago University Press. Include a DOI after the publisher, if available.
-------------------	---

In-text citation	... (Easton, 2008) or according to Easton (2008)
------------------	--

Book & eBook with DOI

In Reference list	8. Ewert, E. W., Mitten, D. S., & Overholt, J. R. (2014). <i>Natural environments and human health</i> . CA International. https://doi.org/10.1079/9781845939199.0000 Include the DOI using the format https://doi.org/10.xxxx/xxxx
-------------------	--

In-text citation	(Ewert et al., 2014) or according to Ewert et al. (2014)
------------------	--

Book & eBook without DOI, eBook without DOI from research databases

In Reference list	9. Foxall, G. R. (2018). <i>Context and cognition in consumer psychology: How perception and emotion guide action</i> . Routledge. Use the copyright date. For more than one publisher, list all publishers in the order that they appear, separated by a semicolon. Do not include the database name or URL.
-------------------	--

In-text citation	(Foxall, 2018) or According to Foxall (2018)
------------------	--

Article with DOI from research databases

In Reference list	10. Washington, E. T. (2014). An overview of cyberbully in higher education. <i>Adult Learning</i> , 26(1), 21-27. https://doi.org/10.1177/1045159514558412 For DOI, use the format https://doi.org/xxxxxx
-------------------	--

In-text citation	(Washington, 2014) or Washington (2014) stated that
------------------	---

Article without DOI: From research databases, or From a Print journal

In Reference list	11. Moody, M. S. (2019). If instructional coaching really works, why isn't it working? <i>Educational Leadership</i> , 77(3), 30-35. Do not include the database name or URL for articles from most academic research databases.
-------------------	--

In-text citation	(Moody, 2019) or Moody (2019)
------------------	-------------------------------

Open access journal article without DOI

In Reference list 12. Dayton, K. J. (2019). Tangled arms: Modernizing and unifying the arm-of-the-state doctrine. The University of Chicago Law Review, 86(6), 1497-1737. <https://bit.ly/2SkWwcy> Include the URL for articles without a DOI and with a nondatabase URL, e.g. free online journal articles

In-text citation (Dayton, 2019) or Dayton (2019)

Webpage, no date

In Reference list 13. Athletics New Zealand. (n.d.). Form a new club. <http://www.athletics.org.nz/Clubs/Starting-a-New-Club> When the author and site name are the same, omit the site name.

In-text citation (Athletics New Zealand, n.d.) or Athletics New Zealand (n.d.)

Webpage, with a date

In Reference list 14. Monaghan, E. (2019, December 10). 5 reasons modern slavery at sea is still possible in 2019. Greenpeace. <https://bit.ly/2PIXjqc> If the page we are citing includes a last updated date, use that date. Include the name of the website in the reference.

In-text citation (Monaghan, 2019) or Monaghan (2019)

YouTube video or another streaming video

In Reference list 15. MSNBC. (2020, January 7). Julián Castro endorses Elizabeth Warren [Video]. YouTube. <https://www.youtube.com/watch?v=Uk2Tzc8H5po>

In-text citation (MSNBC, 2020) or MSNBC (2020)

Conference paper

In Reference list 16. Mason, I., & Missingham, R. (2019, October 21-25). Research libraries, data curation, & workflows [Paper presentation]. eResearch Australasia Conference, Brisbane, QLD, Australia. <https://bit.ly/2RGcFdn> After the title, include a label in square brackets that matches the description of the presentation e.g. [paper presentation]. Create a short URL using <https://bitly.com>

In-text citation ...act of curation works with unique set of factors (Mason & Missingham, 2019) or ...according to Mason and Missingham (2019)

Let us check our progress-2

Discuss the main components of research report.

6.2 Verbal or oral presentation of research reports

It is the delivery of a speech to an audience. It is a face to face about the facts observed or researched. It is a presentation of factual information by mouth or the spoken words. It is usually focused, concise and interesting on result of research topic. It should include a short introduction, objectives, hypotheses, methodology, graphs or tables of findings and interpretation related to findings. It should follow three steps- planning, practicing and presenting. There are four types of oral presentation-

- (i) **Informative** - It is used to give information to concerned audience.
- (ii) **Instructional**- It is used to give instructions or teaching to concerned audience/students.
- (iii) **Arousing**- It is used to arouse interest of the concerned audience.
- (iv) **Persuasive**- It is used to make audience accept and agree on presenter's proposal.

Merit- It provides an opportunity for give and take decisions to a better understanding of findings and their implications.

Demerit-It is the lack of any permanent record concerning the research details. It may have chance of fade away from people's memory even before an action is taken.

Solution-It is effective if supplemented by various visual devices, e.g. use of slides, wall charts and black boards. Distributing a broad outline, with a few important tables and charts related to research results will make listeners attentive.

6.3 Poster presentation

Poster- A poster is a static and visual medium which is used to communicate ideas and messages. A poster is made up of four key characteristics- title, graphic, text and white space. These four key characteristics are influenced by lay out, flow and colour. A poster should contain a title, abstract, short introduction, objectives, methodology, results, discussion and conclusions.

Poster presentation- A poster presentation is a form of presentation of research findings. It is an academic or professional activity. It is shared with conference participants. When many posters are displayed at a time, it is known as poster session. The information is written on posters should clear and concise. The posters are displayed in highly visible areas. Posters may in different sizes but the standard poster size should be in A-1 format i.e. the width of the poster should have 594 mm, length should have 841mm. Posters can be pinned horizontally or vertically. The font size of posters should not too small, because posters will also be made available on conference flash boards.

Poster presentation is an oral presentation. In which, we have more control over our audience can focus on. As we proceed, we can tell a story and lead our listeners along particular paths of reasoning and thought.

Research poster- Posters are widely used in the academic and professional community. Most of the conferences include poster presentations in their programmes. In these conferences, research findings are presented through posters. These posters are called research posters, which contains the summary or concise research information. It attracts

the public and start discussion. Usually, the poster contains a mixture of tables, pictures, graphs etc. In any conference, the researcher/presenter stands by the poster displayed. Other participants can view the presentation and interact with presenter for clarification.

The following points make poster qualitative and effective-

- ❖ The displayed information should be readable from the distance of ten feet.
- ❖ The title should be short as well as attracting the interest of viewers.
- ❖ It should have about 300-800 words.
- ❖ The text should clear and to the point.
- ❖ It should include the use of graphics, colour and fonts.
- ❖ It should have consistent and clean layout.
- ❖ It should have use of bullets, numbering and headlines.
- ❖ It should also have acknowledgements, name of poster presenter and his/her institutional affiliation.

The following points should keep in mind before presentation of poster

- ❖ The poster should have quality of conversation starter
- ❖ We should have to know our audience or participants
- ❖ We should have to know the facts which are going to share with our audience or participants
- ❖ We should make draft of the content which are going to share with our audience or participants
- ❖ We should use the best way of presentation
- ❖ Network should be prepared before starting of presentation
- ❖ Start presentation and engage audience or participants

A good poster presentation should base on a few main ideas. It should not be used more than three different font sizes- the smallest font size for text, the second largest for titles and the largest for the poster title or heading. The colour used in the poster should simple.

Let us check our progress-3

Write short notes on: (a) Oral presentation (b) Poster presentation.

6.4 Writing scientific article/ paper

As a researcher, writing a manuscript may see a good way to promote self. A good research article listed on our resume will always create a good impression. The scientific article or a research article is published in a professional scientific journal. These journals may be of different levels e.g. international, national or local levels. Each journal issues guidelines for submitting articles. The authour has to follow the guidelines in writing and submitting article for a particular journal. The articles may be in form of full article, review article, editorial article, letters and short communications.

Submission of articles

The article of interest may vary from researcher to researcher. The readers may have different preference. But the importance of each article e.g. a full article or review or a short communication is equal. So, before we start writing make ourselves well versed

with the guidelines accordingly. It is very important to choose the right type of article that suits our need and design. The following points may be considered-

- ❖ Visit particular journals webpage. Browse for the guideline, timeline and other necessary information.
- ❖ The proposed article should fall under the scope of the particular journal.
- ❖ The publication option should meet the needs of author.
- ❖ Submit the article to one journal at a time.
- ❖ Communicate with co-authors and give credit them for their contributions.
- ❖ Ask to supervisor or colleagues, if it is first article of the author.
- ❖ Avoid any ambiguous words and check for redundancy.
- ❖ Reading some previously published articles will always help in constructing a good article.

Research articles

If a researcher wants to make their research finding available for others, he/she may publish their work in any Journal. This type of article falls under original research work that has not been published elsewhere. It is ideal for research article.

A research article follows a standard format. It includes abstract, description of research, findings and conclusions that on a researcher has reached. Usually an abstract of a research article contains 150-250 words and full article contains 3500-5000 words according to particular guidelines.

Review articles

Review articles are the focuses on the current understanding of the subject or the work that has been published previously. These articles summarize the previous and current trends and give a reader a sound understanding of the topic. Usually an abstract of a review article contains 150-300 words and full review article contains up to 6000 words according to particular guidelines. It may contain references up to 100.

Popular articles

Focusing on the need of the hour the topic that needs to be elicited to provide a sound understanding of the subject or research work that has been published recently or is gaining interest in recent times comes under popular articles.

Usually an abstract of a review article contains up to 250 words and full review article contains up to 6000 words according to particular guidelines. It may contain references up to 60.

Editorial articles

Editorial articles are generally written by the honorable members of the editorial board of the journal. The editorial articles are depending upon the topic and issue it deals with. Usually it is not more than 1500 words.

Manuscript structure

In any, scientific communication follows a rigid pattern or format that an author should follow.

Here is a general format that an article follows:

S.No.	Section	Purpose
1.	Title	It should reflect the main issues in manuscript.
2.	Authour	The names of all the contributors are written.
3.	Abstract	It is a brief summary of the research and conclusion.
4.	Keywords	These are identifying words which make sure that the article is easily identified.
5.	Background	It is baseline of the research.
6.	Methods	It describes the techniques followed in the research.
7.	Result	These are the main findings of the research.
8.	Conclusion and discussion	These are inferences of findings with justification.
9.	Acknowledgements	These are thanks and gratitude towards people who helped.
10.	References	It is previously published work, should be recognized as well.
11.	Supplements	These are additional information related to the research.

Title

The title is enabling the reader to introduce and decide them if they want to read the article or not. They only read the article, if it is relevant to their field. The title should highlight the content of the manuscript. It should not more than fifty words. It should simple, attractive and able to draw the attention of readers. It should have capacity of stating the content of the article clearly and in a precise manner. It should have unambiguous and specific words. It should avoid unnecessary words and punctuations. Example: Acute Spinal Cord Injury: Assessment and Management

Authours

All the authors who have made a contribution to the research deserve to be credited for the work and the names should be listed in the same line separated by a comma. It is not necessary that the corresponding author should always be the first author and the order of name may vary as per our convenience. But it is advised to highlight the name of the corresponding author with a star (Example: Sher Singh Morodiya*) so as to make it easy for the readers to identify the corresponding author. Correspondence should include email address, telephone and postal address of which email and postal address are mandatory. Now days, the mobile number of the author (s) is also an important need.

Affiliation or address

It is mandatory to include the affiliation or address(s) of the institute where the work has been carried out.

Example: Acute Spinal Cord Injury: Assessment and Management

B.L. Bairwa¹, Sher Singh Morodiya², *

1. Assistant Professor, Sawai Man Singh Medical College, Jaipur, Rajasthan, India

2. Nursing Superintendent, Sawai Man Singh Hospital (SMSMC), Jaipur, Rajasthan, India

Abstract

Abstract includes brief about the introduction, background and a brief account of the methodology adopted, results findings and conclusion of the research. It describes the research study in a fewer words. It contains 150-250 words and avoids the use of tables, figures or any other type of illustrations. References need not be mentioned.

Keywords

Keyword is very important as it will increase the chances of the article to be found. The keywords should reflect the research. It would be an advantage if the keywords are already used in the title. The limits of the keywords should 3-7.

Background

Background should be brief. It highlights the aim of the study or research. It should also be stated the relevant finding and results of other studies. On the basis of challenge and the issues found the article extended.

Methodology

Commonly this section is known as materials and methods. It should include tools and techniques; statistical method or software used the data. The article written in past tense because as the research has already been conducted.

Results

This section should include a detailed account of the recorded findings. The author should provide a clear explanation about the findings, their significance and relevance. Raw data are not included in any scientific article. Raw data are analyzed and should represent it in form of graphs, charts or tables.

Discussion and Conclusion

This section describes the meaning of the research finding. It should also include interpretation of the results and its significance. It should also answer the hypothesis. It should include findings of others' studies. A comparison should make with others' studies. The conclusions should be based on the findings and should have a logical explanation. It should not base on biased judgments.

Acknowledgement

It should include the names of all the contributors that have helped in conducting research.

Supplements

If there is any supplement raw data or any other documented evidences. If it is felt the importance of it, it may include in the supplementary material or can give the URL of the page from where the reader can download the content.

References

Previous research has tested knowledge and already available for public. So, it is important to acknowledge any previously published work.

Citation

The citation of the previously published work appears in the content of the article. It should be mentioned in Vancouver or American Psychological Association format as guidelines of the journal.

Reference list

List of the references as it appears in the text should be included. The number of references should not more than sixty.

Presentation

Section Headings or typography

Typography is one major component of the written form of communication. It is very important to follow the prescribed format.

Headings

There can be three levels of headings

Level 1: Main heading- It should be bold, all capitals and numbered. Mostly font used should be Times New Roman (bold) and 12 points. (It may differ journal to journal).

Level 2: Sub heading- It should be italics, have a space between the main heading and the content. It should also be numbered under the main heading. Mostly font used should be Times New Roman (italics) with 10 points. (It may differ journal to journal).

Level 3: Sub-sub heading- It should be italics and not have space between the content paragraphs. Mostly font used should be Times New Roman (italics) with 10 points. (It may differ journal to journal).

Bullets

Bullets can be used anywhere in the content but size of the bullet should be small.

Line spacing

Line spacing in the content should be 1.0

Colours and highlight

Font and text colour should be black and should not highlighted anywhere in the article.

Images, tables and graphs

Images, tables and graphs should be numbered. It should always be mentioned whenever referencing to them in the article.

Let us check our progress-4

Prepare a general format of article writing.

7. Utilization of research findings

Research utilization is a process of synthesizing, elaborating and using knowledge, which is generated as a result of research. It is used to change in the existing nursing practice. Although the research in nursing was practicing since the time of Florence Nightingale, the pioneer of modern nursing. It was utilized that time; for example, she researched on the health of the British army which was living in India. She reported on bad condition of drainage, water, overcrowding and poor ventilation. These conditions were considered the causes of high rates of deaths on that time. She recommended improving the sanitary conditions of the whole country. This research was utilized and

as a result, it (the drainage and development of sanitary conditions) contributed a significant role in increasing national life expectancy.

Only traditional nursing procedures are carried out also in current days. There is need of research in field of nursing but it is not done at such level of utilization. The nursing personnel are not interested in doing research as well as in utilization of research. They should realize the need of it. Only nursing personnel are not responsible for the utilization of research findings in field of nursing but nurse administrators should take active part.

There is need of institutional support for research utilization. It fosters an environment that expects excellence in practice. The nursing personnel should involve in nursing research activities. It will increase their value role as data collectors. It has been shown to increase research awareness and utilization of findings.

The decisions are carried out at higher level of administration and policy makers. There is gap among areas of education, administration, clinical and research in field of nursing. The following are barriers in utilization of nursing research findings:

- (a) The nursing personnel neither understand the importance nor interested in participating, conducting and implementing the research.
- (b) The nursing personnel have lack of time, confidence, motivation, resources to participate in conducting and implementing research.
- (c) In our country, most of the clinical nursing personnel have diploma as their academic qualification, so they are unaware about the research.
- (d) Neither administration/organization nor nursing personnel are interested or agree to change according to findings of research.
- (e) The nursing personnel are unaware of research because they do not read any research journals.
- (f) The nursing personnel are in confusion of their rights to minor change in routine care.
- (g) Research reports presented by nursing researcher are not communicating to nursing personnel, so that it is not to use new ideas.
- (h) There is a very less component of research in bachelor's programme. The practical component is not carried out in all colleges.
- (i) There are many types of communication problems among researcher, administrator and nursing personnel.
- (j) Our health care system does not stress on the nursing research.
- (k) Administration/ organization do not permit to implement the nursing research.
- (l) The officers who are engaged in planning and implementing the health care system, neither aware nor they are interested in utilization of nursing research.

The following strategies may be used to facilitate utilization of nursing research

The identification of clinically relevant problems is the most important factor in facilitating research utilization. There are multidimensional barriers in it. So, there

is need of a combined and comprehensive effort not from nursing personnel (engaged in education, research, administration and clinical practice) but from ministry of health, directorate of health services, medical colleges and various levels of health care delivery systems also.

Nurse Educators

- (a) Use research finding in teaching-learning activities.
- (b) Include research findings in clinical assignments.
- (c) Motivate students to read research article on the topics which are taught in classes.
- (d) Motivate students to read previously done researches kept in library and think about to conduct research themselves at bachelor's level.

Nurse Researchers

- (a) They should solve the current clinical problems on basis of the research activity.
- (b) They should take research on topics of interest and replicate the results of previous studies.
- (c) They should add strength to findings or disprove beliefs commonly held.
- (d) Elaborate findings of research as early as possible.
- (e) Present research findings at various levels e.g. locally, regionally and nationally.
- (f) Publication of research should be done.

Nurse Administrators

- (a) They themselves know the importance of research.
- (b) They should try to establish research friendly culture in organization.
- (c) They should collect examples from other organizations and try to implement the changes according to research findings.
- (d) Dissemination of research results is important as it may provide options for improvements in practice, communication and better patient outcomes.
- (e) Encourage the nursing personnel to learn implementations of research findings.
- (f) They encourage the nursing personnel by reward for risk taking and innovations.
- (g) They should provide facilities to access research available in digital media.
- (h) Encourage and support continuing education e.g. conferences, seminars etc.

Nurse Clinicians

- (a) They should encourage updating their research skills.
- (b) They should stay updated with literature related to clinical nursing.
- (c) They should engage in continuing education in form of taking higher degrees, specialty specific diplomas, joining professional associations/organizations, pursuing advances courses/degrees.
- (d) They should collaborate with researchers to ask clinical issues and questions.
- (e) They should support conduction of research in the clinical setting.
- (f) They should take the risks to make changes and improve nursing practice.
- (g) They should present evidence via brochures, team meetings, workshops, or social media.

- (h) They have many opportunities for sharing new knowledge. They need not only writing for journals or books, but they also may use social media, presenting conferences and other events related to the research outcomes.

Let us check our progress-5

Narrate the utilization of nursing research.

8. Critical review of published research

Research is a scientific process; therefore, it should be evaluated scientifically. If it is not evaluated scientifically or the implications are not assessed properly, the researcher's hard work will not highlight or acknowledged.

A typical research report includes abstract, an introduction, review of literature, objectives and hypotheses, description of research procedure, findings, references and annexure. Let us take these points one by one to evaluate the research report

(i) Introduction and background- It is first chapter of research report. It says best about the researcher, his/her affiliations and the purpose for which the study was undertaken. If the researcher has a commitment to a particular point of view and supports an assumption or a hypothesis, there is a likelihood of biases. The researcher gets emotionally involved and brings in a bias which is easy to detect. Biases can be detected by seeing researcher's strong view, emotionally charged or indifferent language in the report. He/she may be bias toward socio-cultural affiliation. First point to identify if there are any biases, extent of the bias and its impact on the research process and findings.

The introductory chapter provides a broad overview, indicates the candidate understands in broad spectrum to narrow down to the problem under research.

The entire chapter should be evaluated/ reviewed from several angles

- (a) The broad framework of related problem area lay out.
- (b) The selection of the problem- arguments to justify the study.
- (c) Skills in narrowing down to the identification of the problem
- (d) The presence of biases issues.
- (e) Ability to convince the reader and the reviewer that the problem chosen is relevant, important, timely, researchable and within the competence of the researcher.

If the above criteria be fulfilled, the chapter on introduction should be considered to have been well done.

(ii) Review of literature

Research reports/papers comprise three major elements:

(a) An introduction justifying the research- Some researcher review literature as a part of the introductory chapter to build up the required rationale. A very large number of researchers provide a separate chapter on the review of literature.

(b) Methodological details- As an evaluator of a report, it is to be checked that if the researcher has adequately analyzed and reflected on the previous research studies from the various aspects of research methodology (sample, sampling technique,

sample size, research design, variables under study, scaling techniques, research tool, data collection, techniques for data analysis). Does he/she indicate his/her decision to choose the research design, variables etc. backed up by previous studies? If so, the purpose of review has been well achieved. If not, it is merely a ritual.

(c) Findings and their implications- Major objectives are as follows:

- ❖ **To find gap in research-** If one could conceptually map out a broad area of research and see what researches have been already conducted, the exercise will automatically provide the gaps. Within a broad research area, a number of variables are likely to be impacting a particular phenomenon. Some variables may have been and some others may not have been studied. Those variables which have not been studied are the indicators of gaps.
- ❖ **To find the areas of overlap-** There will be areas in which several studies have been conducted in one and the same way. The same variables have been used time and again in the research studies, which are more or less similar. These are the cases of overlap.
- ❖ **To identify contradictions-** It is quite likely that researches conducted in one and the same area provide contradictory results. The review of literature also brings these contradictions to surface.

As an evaluator, it needs to check out if the researcher has indeed identified gaps, overlaps and contradictions. If so, he/she has done his/her job. It is also to make a qualitative assessment of how effectively the researcher has argued the cases of gaps, overlaps and contradictions.

There is not a single way of reviewing literature. There are four basic patterns:

- (a) First pattern, this is the most elementary pattern where a researcher presents the findings of a study against the name of the author in one paragraph. The second paragraph refers to another, third to yet another researcher and so on. In such a case, the researcher does not interlink one study with another.
- (b) Second pattern, when a reviewer refers to a particular set of findings and provides a few references of those who contributed to that particular finding in brackets. These clusters of studies around a finding but does not compare any two sets of findings or any two sets of researchers.
- (c) Third pattern, when a reviewer describes a phenomenon for investigating various researches as a support.
- (d) Fourth pattern, when a reviewer/ researcher develops a conceptual framework of his/her research in a particular field. Within the framework, the researcher maps out the previous research. Fitting it into a conceptual framework allows him/ her to compare and contrast issues and findings, identify the gaps, overlaps, contradictions and also derive methodological implications.

The above four patterns are in taxonomic structure, as an evaluator, we will award higher credit to the researcher who uses fourth pattern and at least to the

one who resorts to pattern one for reviewing research literature.

(iii) Objectives and hypotheses

All the research studies have objectives and hypotheses. It is important to examine if the researcher has raised very clearly the questions to which he/she is looking for a solution. These questions should be clear and categorically put down on paper. This set of questions can be converted into objectives.

The major attributes of well written objectives are:

- (a) Clarity of expression and direction-** The objectives must be clear to indicate what the researcher is trying to investigate. It is equally important to avoid overlaps in stating objectives.
- (b) Measurability-** The objectives must be stated in such a manner that they are measurable. In qualitative research it should be codified the data and information so that assessment can be made if the objectives have been achieved or not.
- (c) Comprehensiveness-** It should be comprehensive enough to cover each and every aspect of the research study.
- (d) Judiciousness-** It is important attribute justifiability of choosing and stating objectives.

While evaluating the research objectives needs to examine clarity of expression, measurability, comprehensiveness of objectives and judiciousness in choosing and stating them. A hypothesis is a statement of casual or non-casual relationship of two or more variables under study. The statement of a hypothesis is derived from theoretical constructions, previous research and logical analysis.

The evaluator has to check if the researcher has provided sound back-up from previous research findings and important theoretical analyses to justify his/her formulation of hypotheses.

Hypotheses are stated either in null or directional form. Null form, does not presuppose any specific relationship. Directional hypotheses presuppose relationship. When a null hypothesis is tested, it may point to a positive, neutral or negative relationship that can be used to derive conclusion. When a directional hypothesis is tested, it may point one of the two results- true or false. If it is false, it does not automatically show that the reverse is true.

While evaluating the research hypotheses, it needs to examine if the choice of hypotheses-null or directional, was logical and if the researcher has adequately argued out his/her case.

Are the hypotheses testable?

The hypotheses are stated clearly indicating one to one relationship between two or more variables.

In case of a multi-variate situation, the relationship of cluster of independent variables related to the criterion variable is well defined.

(iv) Research design

There are several research methods and designs that can be chosen by a researcher to achieve the objectives and test the hypotheses. It is important to evaluate if the chosen design is competent to respond to the research objectives and questions laid down.

In evaluating research study, it is necessary:

- (a) To check the choice of appropriate research design against the objectives.
- (b) Evaluating the applicability and be fittingness against the hypothesis. If hypothesis to be tested is formulated in terms of relationship, the study has to adopt a methodology by which relationships can be tested.

The evaluator of a research report needs to examine the appropriateness of the choice of research design related to the research objectives. The details of the design too have to be evaluated. Equally important is the argument put forward by the researcher in deciding the research design.

(v) Variables

The following points to be considered while evaluating research variables:

- (a) How minutely the variables have been identified and classified under three categories-independent, dependent and intervening.
- (b) If the researcher has provided operational definitions of at least such variables as do not have a standard meaning in the literature.
- (c) If there are clear indications of the measurability of variables.

(vi) Research tool

The following points to be considered while evaluating research tool

- (a) If the researcher has chosen tool that can actually measure the variables.
- (b) If the tool has been picked up from an existing stock. The researcher has to check its validity, reliability and the sample on which the original study was conducted. If the standardization on the original sample is valid, for sample on which the researcher has used for the tool and drawn inferences. In case the tool has been constructed by the researcher, has been taken to check the attributes of the tool, a dependable research tool, be it a questionnaire or an interview schedule.
- (c) If the researcher has tested the feasibility of the use of tool, e.g. a questionnaire not a feasible tool for the illiterates unless the researcher himself/ herself records the responses.
- (d) It should be examined from the angles of language, communication, provision of recording response by the respondents etc.

(vii) Sample

There are following major issues with respect to sample

- (a) Sample size- Size of sample depends on the nature of objectives of a research study and research design, e.g. in case of experimental research, it is difficult to handle large samples. In survey research, samples should be large. There has to be an optimum size of the sample. Here, it is to be considered that the sample size is large enough for the study and the sample size has been determined sci-

entifically.

- (b) Sample technique and type of the sample- There are several techniques of choosing a sample. A random sample is often considered ideal sample, it is necessary for generalization and the creation of a new knowledge. It is important to check, if the researcher has identified the right and relevant criteria for stratification and sub-stratification of the population, then developed a sample frame to choose a stratified random sample.

It is to see the argument given by the researcher justifying the method of selecting of the sample. An evaluator needs to carefully evaluate the explanation provided by the researcher and the way he/she proposes to cope with it.

(viii) Data collection and analysis

The quality of data is determined by the data collection procedure. The quality of data lies in the dependability of information collected from the sample. It is a worst situation, when researcher manipulates the responses by filling research tool himself/herself. It is unprofessional research behavior. On spot collection that provides a return of responses, but not necessarily quality responses because the respondent is likely to respond mechanically. When researcher sends tool by mail, there is large amount of loss. The data generated are not the predetermined sample but of positive type of respondents in the sample. Generally, in a research exercise that requires primary data, it should collect personally. In case of secondary data, it is important to check the sources of data and their trustworthiness. Here it is to check, if the researcher has categorically recorded the detail of data collection

- (a) Research tool administered personally or by a representative or by mail.
- (b) Sources authenticity of secondary data.
- (c) The kind of problems that the data might have in terms of quality.

Data analysis can be either qualitative or quantitative. A main point in evaluating research is to check if the researcher has chosen qualitative methods where the data are qualitative and objectives and hypotheses do not demand a quantitative analysis. Like this, if the researcher has chosen a quantitative technique where a qualitative answer is not required.

In quantitative research the statistics can be parametric or non-parametric.

- (a) It is to check if choice is correct or not.
- (b) The size of sample is another determinant in choosing the statistical tests as a small sample often use in non-parametric tests.
- (c) If the researcher has chosen the right test then size of sample is another important point to consider, e.g. within the application of simple central tendencies, if the researcher chosen mean where a median or mode would have been the more appropriate choice.
- (d) The correctness of calculations. With the increasing use of computers, this problem has reduced.

(ix) Findings and implications

The value of a research study lies in its findings. The findings are presented in the descriptive form as well as in form of tables and graphs. The evaluator needs to examine if a table or a graph has been made, is required. If the table and graph are given a title and explanation. It should also be checked if researcher has made table and graph on same thing, ensure that there is no contradiction between the two. It is necessary to provide an interpretation and implications of the result. Researcher is expected to reflect previous studies that contradict his/her findings. It is to be critically examining the way a researcher interprets the findings

- (a) If the interpretation has been related to the introductory chapter where the researcher builds up the rationale for his/ her research.
- (b) If the argument builds up in each chapter is adequately reflected in the interpretation of the results.

The evaluator examines the skill of analytical and exploration. He/ she examines the skill of observing and explaining relationships between two or more variables etc.

(x) Summary and conclusions

It provides brief look of the entire research study. An evaluator examines how a researcher has prepared the rationale and documented the objectives, hypotheses, research methodology and findings.

(xi) Referencing

It is an important skill. Most research students have lack of skill and the seriousness of purpose. Many evaluators check the reference section because there are some definite internationally recognized formats. Many students do not know the difference between a bibliography and references. Research reports needs references not a bibliography.

- (a) It is necessary to check if all the information has been provided is in one of the standard formats.
- (b) It has to be seen if the references are indexed, but they are not numbered in the text and vice-versa. An evaluator goes through the text of the report and comes across a reference in the text, he/she should immediately check if a corresponding entry in the reference. It identifies the missing references and common mistakes in the spellings of names and years of the publications.

(xii) Annexure

The main purpose of the annexure is that an evaluator can check the actual quality of the research tool and material that have been used by the researcher. It needs to be fully documented and also serialized as indicated in the content of the research report. The points for evaluations are the comprehensiveness and serialization of the annexure.

(xiii) Other indications

The following are some indications which also determine the quality of a research report, are

- (a) Language and expression (syntax, spelling etc.)

- (b) Typing, word processing and printing of report (readability, margins, line spacing, type font and size, placement of tables, diagrams, illustrations and graphs)
- (c) Binding and overall setup.

9. Let us sum up

When procedure of data collection has been completed, there is need of analysis and interpretation of data. The finding of the research study is communicated by different methods. These findings may also be reviewed/ evaluated by other researchers, funding agencies etc. These are utilized by concern persons/organizations/authorities. Tabulation is the process of placing classified data in tabular form. Compilation is a process of condensing information by classifying and tabulating vital statistical data into various categories or groups with the object of producing vital statistics according to a determined tabulation programme. Summarization is a process of taking a lot of information and creating a condensed version that covers the main points. It is a short conclusion of a big theory or a paragraph. Data analysis includes categorizing, ordering, manipulation and breaking down of data, which helps to test the hypotheses and to obtain the answer of research question. It is based on the research objectives. There are two types of analysis- descriptive analysis and inferential analysis. Interpretation of data is a task of drawing inferences from the collected data after research study. It helps researchers to categorize, manipulate, and summarize the information in order to answer critical questions. There are some steps involved in technique of interpretation. All the activities of research go waste unless the findings are recorded and reported to concern persons. A research report is a precise presentation of the work done by a researcher on a particular problem. The reporting of research depends on the purpose with which it was undertaken. The content of the research report never varies. Style of writing report may differ according to purposes. Verbal or oral presentation of research reports is a presentation of factual information by mouth or the spoken words. It should include a short introduction, objectives, hypotheses, methodology, graphs or tables of findings and interpretation related to findings. It should follow three steps- planning, practicing and presenting. A poster is a static and visual medium which is used to communicate ideas and messages. A poster presentation is a form of presentation of research findings. When many posters are displayed at a time, it is known as poster session. Research poster contains the concise research information. It attracts the public and start discussion. Usually, the poster contains a mixture of tables, pictures, graphs etc. In any conference, the researcher/presenter stands by the poster displayed. Other participants can view the presentation and interact with presenter for clarification. The scientific article or a research article is published in a professional scientific journal. These journals may be of different levels e.g. international, national or local levels. Each journal issues guidelines for submitting articles. The authour has to follow the guidelines in writing and submitting article for a particular journal. The articles may be in form of full article, review article, editorial article, letters and short communications. A research article is the orig-

inal research work that has not been published elsewhere is ideal for research article. It includes abstract, description of research, findings and conclusions that on a researcher has reached. Usually an abstract of a research article contains 150-250 words and full article contains 3500-5000 words according to particular guidelines. Review articles are the focuses on the current understanding of the subject, or the work that has been published previously, these articles summarizes the previous and current trends and gives a reader a sound understanding of the topic. Editorial articles are generally written by the honorable members of the editorial board of the journal. Usually it is not more than 1500 words. In any, scientific communication follows a rigid pattern or format that an author should follow. Research utilization is a process of synthesizing, elaborating and using knowledge, which is generated as a result of research. It is used to change in the existing nursing practice. Although the research in nursing was practicing since the time of Florence Nightingale, the pioneer of modern nursing. There is need of research in field of nursing but it is not done at such level of utilization. There is gap among areas of education, administration, clinical and research in field of nursing. There are multidimensional barriers in it. So, there is need of a combined and comprehensive effort not from nursing personnel (engaged in education, research, administration and clinical practice) but from ministry of health, directorate of health services, medical colleges and various levels of health care delivery systems also. Research is a scientific process; therefore, it should be evaluated scientifically. If it is not evaluated scientifically, neither the implications be assessed properly nor the researcher's hard work will remain acknowledged. In the next chapter, we will introduce to the statistics.

10. Answers to check our progress

Let us check our progress-1

Write the meaning of followings:

- (a) Tabulation - Please refer para 3.1, page 71.
- (b) Compilation- Please refer para 3.2, page 71.
- (c) Summarization- Please refer para 3.3, page 71.

Let us check our progress-2

Discuss the main components of research report.

Please refer para 6.1, page 73.

Let us check our progress-3

Write short notes on:

- (a) Oral presentation- Please refer para 6.2, page 95.
- (b) Poster presentation- Please refer para 6.3, page 95.

Let us check our progress-4

Prepare a general format of article writing.

Please refer para 6.4, page 96.

Let us check our progress-5

Narrate the utilization of nursing research.

Please refer para 7, page 100.

CHAPTER- 6

Introduction to Statistics

Structare

1. Introduction
2. Objectives
3. Statistics
 - 3.1 Basic terms
 - 3.2 Definitions
 - 3.3 Characteristics
 - 3.4 Limitations
 - 3.5 Use of Statistics
 - 3.6 Importance of statistics
 - 3.7 Types of statistics
4. Measurements in statistics
 - 4.1 Data
 - 4.2 Continuous and discrete series
 - 4.3 Scale/ levels of measurement
 - 4.4 Determining the meaning of test score in a continuous series
5. Descriptive statistics
 - 5.1 The frequency distribution
 - 5.2 Graphic presentation of the frequency distribution
 - 5.3 Measures of central tendency
 - 5.4 Measures of variability
6. Normal probability
7. Test of significance
 - 7.1 Meaning
 - 7.2 p-values
 - 7.3 The significance levels
 - 7.4 The level of confidence
 - 7.5 Types of errors
8. Co-efficient of correlation
 - 8.1 Meaning
 - 8.2 Types and degree of co-relation
 - 8.3 Methods of determining co-relation
9. Statistical packages and its application
 - 9.1 Statistical packages meaning
 - 9.2 Application of statistical packages
10. Let us sum up
11. Answers to check our progress

1. Introduction

The subject of statistics is not a new discipline but it is old as the human society itself. The word "statistics" was first used by German mathematician Gottfried Achenwall in 1749, who is also known as the "father of statistics". We may say that the word "statistics" has been derived from the Latin word "Status" or Italian word "Statistia" or German word "Statistik" or the French word "Statistique". It means a political state or information useful to the state. It appeared in the encyclopedia of Britannica in 1797. It was used by Sir John Sinclair in Britain. In the 19th century, the word "statistics" acquired a wider meaning. It covers numerical data of any subject. Now, it is being used in different meanings.

- ❖ It is defined as a "discipline" that includes procedures and techniques used to collect, process and analyze the numerical data to make inferences. Here, it is always used in a singular sense.
- ❖ It refers to "numerical facts" that are arranged in the form of tables or charts etc. Here, it is always used a plural, e.g. deaths, accidents, crimes, births, nursing colleges etc.
- ❖ It denotes that statistics are "numerical quantities" calculated from sample observations. A single quantity calculated from sample observations is called statistics such as the mean. Here word statistics is plural.

In this chapter, we will define statistics and explain the uses of statistics. We will describe the scale/levels of measurement, frequency distribution and graphical presentation of data. We will know the calculation of mean, median, mode and standard deviation. We will discuss normal probability, test of significance, co-efficient of correlation. We will explain the statistical packages and its application.

2. Objectives

After studying this chapter, we will be able to:

1. Define statistics.
2. Describe the scale/levels of measurement.
3. Tell the uses and types of statistics.
4. Know about frequency distribution and graphical presentation of the data.
5. Describe and calculate the mean, median, mode and standard deviation of data.
6. Understand the normal probability.
7. Explain the test of significance.
8. Discuss about the co-efficient of correlation.
9. Describe statistical packages and its application

3. Statistics:

3.1 Basic terms

Following are the basic terms used in statistics:

Data: Data are individual pieces of information recorded during research study. These are used for the purpose of analysis.

Qualitative data: Data, which are demonstrated or available in form of detailed description of events, situations, people, interactions and observed behavior, direct quotations/statements from people about their experiences, beliefs, attitudes, thoughts and a short extract from records, correspondence, documents and case histories.

Quantitative data: Data, which are communicated in form of nominal, ordinal, interval or ratio scales of measurement.

Discrete data: These data has a clear space between values. These can take on certain values. These values do not have to be complete numbers but they are values that are fixed. It is only containing finite values, the subdivision of which is not possible.

Continuous data: These data can be measured on an infinite scale; it can take any value between two numbers, no matter how small.

Parametric data: These are gained by applying interval or ratio scales of measurement.

Non-parametric data: These are obtained by applying nominal or ordinal scales of measurement. These are either counted or ranked.

Central tendency: It provides a single most typical value as representative of a group of values. The trend of a group of measures as indicated by some type of averages, usually the mean, median or mode.

Mean: It is a kind of average obtained by dividing the sum of a set of measures by their number.

Median: It is the middle value in a distribution or set of ranked values; the point that divides the group into two equal parts.

Mode: It is the value that occurs most frequently in a distribution.

Variability: It is the dispersion or spread of measures or values.

Range: It is the difference between the highest and the lowest obtained measure or value on a tool, for some groups. It is rough measure of variability.

Variance: It is a measure of variability of a distribution. It is the average of the squared deviations of the measures or values from the mean.

Standard deviation: It is the positive square root of variance.

Normal distribution: It is a distribution of measures in a graphic form. It has a distinctive bell-shaped appearance. It is a symmetrical and asymptotic. The mean, mode and median for this type of distribution have equal values.

Percentile rank: It is the expression of an obtained test score in terms of its position within a group of 100 scores.

Co-efficient of correlation: It is a measure of the degree of relationship between two sets of measures for the same group of individuals. Its value ranges from 00 (denoting a complete absence of relationship) to (indicating perfect positive relationship) +1.00 and -1.00 (indicating perfect negative relationship).

Parameters: Parameters are numbers that describe the properties of entire populations. It tells something about the whole population. (e. g. mean, median or mode of population).

Statistics: Statistics are numbers that describe the properties of samples. (e. g. mean, median or mode of samples).

Parametric tests: These are statistical tests which are used for analyzing parametric data and making inferences about the parameters from the statistics. These tests are based upon certain assumptions about the nature of data distributions and the types of measure used.

Nonparametric tests: These are statistical tests which are used for analyzing non-parametric data and making possible useful inferences without any assumptions about the nature of data distributions.

Standard errors of mean: It is the standard deviation of a distribution of sample means.

Degree of freedom: The number of degrees of freedom in a distribution is the number of observations or values that are independent of each other and cannot be deduced from each other.

Reliability- It is the degree to which a measuring tool/ instrument gives the same results each time that it is used.

Validity- It is the degree to which the tool/ instrument measures, what it is intended to measure.

3.2 Definition

(i) Statistics is “Science and arts that gathers and co-ordinates numerous facts within a determined field, treats these mathematically so that the numerical relations between these facts may be displayed clearly and freed from anomalies and chance factors.”
- *English and English*

(ii) “Statistics comprises the collection, tabulation, presentation and analysis of an aggregate of facts collected in a methodical manner, without bias and related to a predetermined purpose.”
- *W.G. Sutcliffe*

(iii) “Statistics is that branch of scientific method which deals with data obtained by counting of measuring properties of population of natural phenomena.” - *M.G. Kendall*

(iv) “Statistics is the science which deals with the collection, classification and tabulation of numerical facts as the basis for explanation, description and comparison of phenomena.”
- *Lovitt*

(v) “Statistics is that branch of scientific methodology. It deals with collection, classification, description and interpretation of data obtained by the conduct of surveys and experiments. Its essential purpose is to describe and draw inferences about the numerical properties of populations.”
- *George A. Ferguson*

Statistics is considered as a discipline that concerns the collection, organization, analysis, interpretation and presentation of data.

3.3 Characteristics

(a) Statistics consists aggregates of facts

A single and isolated figure or fact cannot be called as statistics. The data (plural sense) are subject matter of statistics not datum. The statistics must consist of

aggregate of certain facts. A single and isolated fact or figure like, 160-centimetre height of a nursing student or the 60 marks obtained in applied anatomy of a particular nursing student in mid-term examination are not count as statistics, because these figures are unrelated and incomparable. For a data may amount to statistics it must be in the form of a set or aggregate of certain facts, e.g. height of 30 students of first year of B.Sc. nursing and marks obtained in applied anatomy of 30 students of first year of B.Sc. nursing in mid-term examination may be count as statistics.

(b) It is affected by multiple causes

It is not easy to study the effects of one factor only by ignoring the effects of other factors in statistics. We have to study the effects of all the factors separately as well as collectively on the phenomenon because effects of the factors can change with change of place, time or situation, e.g. we can say that result of B.Sc. nursing final year examination does not depend on any single factor but collectively on teachers (standard of teachers, teaching methods and teaching aids), students (performance of students in theory, clinical and practical) and other factors (standard of question papers, methods and standards evaluation).

(c) It should be expressed numerically

There is a main characteristic of statistics that the facts be expressed in numbers, so that statistical processes may carry out. The qualitative expression like he is a young man, he is poor/rich or illiterate etc. are not considered as statistics but he is 25 years old and his income is Rs. 30000 per month etc. Thus, these data contain numerical figures so that may be called as numerical statement of facts.

(d) It must be enumerated or estimated accurately

In a particular enquiry, numerical data may be obtained by complete enumeration or by estimation. The use of these both methods depends upon the area and object of enquiry. If the area of enquiry is not very large and complete enumeration is possible, the data collection is made through complete enumeration. On the other side, if the area of enquiry is very large and complete enumeration is not possible or impractical due to certain reasons, the collection of data be made through estimation. It is done by the technique of sampling, e.g. to find out the knowledge level on nursing care of a patient with covid-19 among nursing students in country. In such case, complete enumeration is not possible or impractical due to certain reasons, the collection of data be made through estimation. But if this enquiry is made on a particular nursing institution, it is possible to obtain data through complete enumeration.

(e) It should be collected in a systemic manner

There should be made a suitable and systematic plan, before collecting the data. The trained investigators should be used to collect data. The data collected in a haphazard manner will lead to difficulties in the process of analysis and show wrong conclusions. If, it is not done, in such cases reliability of data gets decreased. So, to get

correct results the data must be collected in a precise manner.

(f) It should be collected for a predetermined purpose

We must be clear about the purpose and scope of enquiry before we start the collection of data, if we have no information about its purpose, we may not be able to collect data according to the needs. The definitions of various terms used, unit and measurement help to achieve the required purpose. Suppose we want to get data on preparedness of vaccination against the covid-19 virus. We have to know about total community health centres, primary centres, sub centres along with available manpower which may be utilized for the predetermined purpose.

(g) It should have a reasonable standard of accuracy

The estimation of facts and figures cannot be as precise and accurate as actual enumeration or actual measurement. A high degree of accuracy observed in mathematics is not observed in statistics, e.g. while measuring the weight of a patient accuracy will in terms of a fraction of a gram, while measuring the weight of loaded truck accuracy will in terms of a fraction of a kilogram. "What should be the degree of accuracy depends upon the object of enquiry." However, certain standards of accuracy be maintained for drawing meaningful conclusion.

(h) It should be capable of being placed in relation to each other

In any research study, the collection of data is generally done to compare. If the figures collected are not comparable, in that case, they lose a large part of their significance. The comparison is done between homogeneous study and not heterogeneous, e.g. heterogeneous data like average weight of B.Sc. nursing first year students 54 kilograms and distance between college of nursing and clinical area is 10 kilometres can never be placed in relation to each other and compared for analysis and interpretation. It can be concluded that all statistics are numerical data but all numerical data are not statistics unless they satisfy all the essential characteristics of statistics, described above.

3.4 Limitations

There are so many applications of statistics in every field of life. But it is not free from limitations. The following are main limitations of statistics

(a) Statistics deals with aggregates of facts not with individuals

It always studies a group of values instead of single observation studies, e.g. the study of marks of one student of a nursing college is not statistics, but the study of marks of 50 students of that nursing college form statistics. It can be said in other words, the marks of 50 students do not constitute statistics but the average marks of 50 students of a nursing college forms statistics.

(b) Statistics is able to study only quantitative characters not qualitative

Generally, it studies only the quantitative characters of the given problems instead of qualitative characters. The problems which cannot be studied quantitatively (i.e. in numerical form) such as intelligence poverty, honesty etc., are not directly studied in statistics. If we want to studies these problems, we need to change in quanti-

tative form, e.g. intelligence of any nursing student may be studied on the basis of the marks secured by him in a particular examination.

(c) Statistical laws are not accurate

Laws of natural and physical sciences are based on exactness but statistical results are almost not exactly correct. In statistics, we talk on the basis of probability and not certainty. A hundred percent accuracy is impossible in statistical work because statistical laws are based on average.

(c) Statistics deals with averages

It deals with averages that are obtained from different individual items. The statistical laws are true only on average, e.g. average marks of nursing student cannot be used for forming idea about marks of an individual student. It should be studied the dispersion of marks which show the degree of variability in marks of the units.

(d) Statistics can be misused

When the statistical methods are used by incompetent, unskilled and inexperienced persons, it may lead to false conclusions. Only the expert statistician can scientifically handle statistical data.

(e) Statistics is only one method to study a given problem

It is a method of studying a problem. It is not best solution for all conditions of the given problem. It cannot be much helpful in studying the provided problem, e.g. cultural, religious and philosophical unless they are supplemented by other evidence.

(f) Statistics is used with a given purpose

The validity of a statistical result in a particular matter, it is not necessary that it would be utilized in another matter. The use of secondary data of other study should be used carefully. In large number of cases, statistical data are based on sample study. The various samples got from the same universe may give varied results, so applying the result on the whole universe may dangerous.

(g) There are too many methods to study problems

It uses so many methods to find a single result. Variation can be found by quartile deviation, mean deviation or standard deviations and results vary in each case. It must not be assumed that the statistics are the only method to use in research, neither should this method of considered the best attack for the problem.

(h) Statistics does not reflect the entire story

It only gives solutions of certain quantitative problems. The real background of the data may not be revealed in these facts.

3.5 Use of Statistics

The knowledge of statistics helps us in utilizing the proper methods to collect the data, the correct analyses and present the results effectively. Statistics is an important process to make discoveries, make decisions and make predictions based on data. It helps in providing a better understanding and accurate description of natural phenomena. It helps in the proper and efficient planning of a statistical inquiry in any field of study.

It helps in collecting appropriate quantitative data.

3.5.1 General use of statistics in our daily life

Statistics is used in every aspect of life, e.g. in data science, robotics, business, sports, weather forecasting and much more. Apart from this, there are various other fields where statistics concepts are applied like running, calculating the budget, calories etc. Statistics can be used in comparing the values, testing hypotheses, weather forecasting etc. It can help in improving the reasoning skills that are necessary to calculate uncertainty. Following are important examples of use of statistics in various areas:

(a) Government- Statistics is utilized by government in making judgments related to health, education, populations etc. There is current example of pandemic i.e. corona novel virus. It is used to check the assertiveness of vaccine, the progressive reports after vaccination, usefulness of vaccines, target and availability of vaccines etc.

(b) Weather forecast- Statistics play a crucial role in weather forecasting. The use of computers in weather forecasting is based on statistical functions. These functions compare the weather condition with the pre-recorded seasons and conditions. The concept of probability and statistics are used by the predictors.

(c) Emergency preparedness- Statistics is used to predict any natural disaster that may happen shortly and also announces which places, they can affect. It helps us to get prepared for an emergency. It also helps the rescue team to do the preparation to rescue the life of the people who are in danger. Statistics helps to the Indian meteorological department to alerts the citizens before it happens.

(d) Political campaign- Statistics are important in any political campaign. Without statistics, no one can run a political campaign perfectly. It helps the politicians to have an idea about chances to win an election in a particular area. It also helps the news channel to predict the winner of the election. It also helps the political parties to know their support in a particular voting zone. It helps the country to predict the future government.

(e) Planning- Statistics plays an important in planning. Without statistics, the planning cannot be possible. It helps to do planning in all levels and areas of life, e.g. business, economics, even personal level etc.

(f) Economics - Statistics and economics are interrelated with each other. Almost every branch of economics uses statistics, i.e., consumption, production, distribution, public finance. Statistics is used for comparison, presentation, interpretation and so on.

3.5.2 Use in field of nursing

Human beings always respond in different manner on issues related to various risk factors, diseases and treatments. So, there is need that all nursing personnel should be able to summarize, interpret and display findings in the right ways.

(a) Enhancing the patient care- Statistics enables nursing personnel to prioritize treatments, determine the needs of a patient and immediate medical attention or fol-

low-up care. The nursing personnel use statistics to identify specific patterns in signs and symptoms. It helps to be ready to any medical changes to patient. It helps to carried out treatment on time by maintaining frequency charts or data sheets to document the timing of medication given to patients.

(b) Supporting evidence-based nursing practices- In current scenario, the nursing profession is going to establish empirical evidence for patient care. For this purpose, there is need that nursing personnel should understand the basics of statistics so that they may be able to read, comprehend and interpret relevant literature.

(c) Comparing the options available for nursing practice- Nursing personnel use statistics to determine the effectiveness of the options of the nursing interventions of their choice. In some settings, nursing personnel use their knowledge related to statistics, when have to make several decisions during critical care.

(d) Healthcare/ nursing care has become more complex and challenging- Nursing has become more challenging due to electronic information and published literature that exist. There is always a need of new knowledge to improve patient care. The nursing personnel have to continue learning during the practice.

(e) There is demand of knowledgeable and competent leadership- Formal nursing education is an important component of the commitment to become a leader in the nursing profession. For one to become a nursing leader, they must be able to keep up with the current trends, be able to solve nursing related problems and know about new technology that can improve patient care.

3.6 Importance of statistics

- (a) Statistics plays an important role in the research.
- (b) It helps to answer the research questions in the field of study.
- (c) It helps to understand the process of application of statistical methods.
- (d) It says about the suitability of a particular research study.
- (e) It enables to understand specified statistical concepts and procedures.
- (f) It involves the collection of data, organizing, analyzing, interpreting and presenting it for drawing conclusions.
- (g) It uses various quantitative tools for gathering and evaluating large sets of data.
- (h) It is an efficient tool that helps businesses in making efficient and well-informed decisions.
- (i) It is used in various fields such as economics, science, medicine, marketing, psychology and politics.
- (j) It helps in converting the obtained raw data into such information that is helpful for developing a better understanding and sound decision making.
- (k) Statistical tools are able to make a better plan by making accurate assumptions and predictions in business.
- (l) They are able to handle future uncertainties in a better way by predicting cyclical and general economic fluctuations.

3.7 Types of statistics

There are two types of statistics- descriptive statistics and inferential statistics.

3.7.1 Descriptive statistics

In this type of statistics, the data is described, summarized and explained. Descriptive statistics are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures. They form basis of every quantitative analysis of data with the help of simple graphics analysis. It presents quantitative descriptions in a manageable form. In a research study, there may have lot of measures. It helps us to convert large amounts of data in a meaningful small form. It reduces lots of data into a simpler summary. It describes what is or what the data shows.

We may take an example of grading system of evaluation of students in many universities, where grading is done as A, B, C, D and E. This single number (A or B or C or D or E) describes the performance of a student of course experiences in that subject/ class standard. The grading system of evaluation tells neither about difficulty nor easiness of course nor their elective or compulsory courses. Besides these limitations, descriptive statistics provide a summary that may enable comparisons across population or other units.

There are some commonly used measures to describe a data set are- central tendency and measures of variability or dispersion. The measures of central tendency include the mean, median and mode. The measures of variability include the standard deviation or variance, the minimum and maximum value of the variables, kurtosis and skewness.

It is a way of organizing, representing and explaining a set of data by the use of charts, graphs and summary measures. There are some common measures which are used to summarize data and present it in tables or graphs, e.g. histograms, pie charts, bars and scatter plots.

3.7.2 Inferential statistics

We use the inferential statistics to interpret the meaning of descriptive statistics. Firstly, the data are collected, evaluated and summarized, then it is used to convey the meaning of the collected data. The probability principle is used in inferential statistics. Inferential statistics are used to test hypotheses and the study correlations between variables. They can also be used to predict population sizes. Inferential statistics are used to derive conclusions and inferences from samples, i.e. to create accurate generalizations.

Let us check our progress-1

- (i) *Define statistics.*
- (ii) *State the characteristics of statistics.*
- (iii) *Write the uses of statistics in field of nursing.*
- (iv) *Enumerate the type of statistics.*

4. Measurements in statistics

4.1 Data

Data are recorded factual information. It is raw information from which statistics are created. It is used for the purpose of analysis, interpretation and presentation in statistics. It is classified into qualitative data and quantitative data. Further, the qualitative data is sub-classified into nominal and ordinal data. On the other hand, quantitative data is sub-classified as discrete and continuous data.

4.2 Continuous and discrete series

4.2.1 Continuous series of data- It is described as an unbroken set/series of observations. It can be measured on a scale. It can take any numeric value, within a finite or infinite range of possible value. The continuous data can be broken down into subdivisions (fractions and decimal) in smaller units, i.e. it can be meaningfully subdivided into smaller parts according to the measurement precision, e.g. dose of a medicine, age of a child, height or weight of a patient etc.

4.2.2 Discrete series of data- The discrete series of data cannot be broken down into subdivisions (fractions and decimal) in smaller units. The discrete series shows real gaps. It is also known as the ungrouped frequency distribution.

Table 6.1 The Difference between Continuous and Discrete Series

S.No.	Base of Comparison	Continuous Series	Discrete Series
1.	Meaning	It falls on a continuous sequence/series. It has not clear spaces between values.	It does not falls on a continuous sequence/series. It has clear spaces between values.
2.	Nature	It is measurable.	It is countable.
3.	Values	It can take any value, in some value within range.	It can take only distinct or separate values.
4.	Graphical Representation	It shows histogram in graphical representation.	It shows bar graph in graphical representation.
5.	Tabulation	Its table shows grouped frequency distribution.	Its table shows ungrouped frequency distribution.
6.	Classification	It contains overlapping or mutually exclusive classifications, e.g. 5-10, 10-15, 15-20...etc.	It contains non-overlapping or mutually inclusive classifications e.g. 5-9, 10-14, 15-19...etc.
7.	Function graph	It shows connected points.	It shows isolated points.
8.	Example	Market price of a product.	Days of the week.

It is clear that continuous data has any value from a given range (without breaks) whereas the discrete data has a certain number of isolated values.

4.3 Scale/ levels of measurement

According to its characteristics, the scale/ levels of measurement are classified in following levels-

(i) Nominal level data- It is first level of data. It is assigned only name and frequency of each category is counted. There is no statistically relationship. Only mode is calculated as measurement of central tendency. When two categories of data occur in the same frequency, it is called bi-modal. When more than two categories of data occur in the same frequency, it is called multi-modal. It is used in those situations where counting is the only feasible method of quantification.

(ii) Ordinal level data- It is the second level of data. It can assign category and rank order, e.g. high, medium, low or from highest to lowest etc. Ordinal level data has no absolute value. When ranking is done, it says equally rank spaces but value wise they are not equally spaced.

(iii) Interval level data- It is the third level of data. It has no absolute zero and therefore no fixed beginning point. It has category, rank and equal spaces or intervals or relative distance between the units and a rule of order. It measures mean as central tendency.

(iv) Ratio level data- It is the fourth level of data. It has category, rank and equal spaces or intervals or relative distance between the units and a rule of order. Besides it, it has absolute zero and therefore fixed beginning point. It is most precise and highest level of data. It has the numerical values are real number and can be added, subtracted, multiplied, divided and expressed in ratio relationship.

Table 6.2 Scale/ Levels of Measurement

S.No.	Types of Data	Mathematical Operations	Measures of Central Tendency	Measures of Variability
1.	Nominal	Equality ($=$, \neq)	Mode	None
2.	Ordinal	Equality ($=$, \neq) Comparison ($>$, $<$)	Mode Median	Range Inter-quartile range
3.	Interval	Equality ($=$, \neq) Comparison ($>$, $<$) Addition (+), Subtraction (-)	Mode Median Arithmetic mean	Range Inter-quartile range Standard deviation Variance
4.	Ratio	Equality ($=$, \neq) Comparison ($>$, $<$) Addition (+), Subtraction (-) Multiplication (\times) Division (\div)	Mode Median Arithmetic mean Geometric mean	Range Inter-quartile range Standard deviation Variance Relative standard deviation

4.4 Determining the meaning of test score in a continuous series

In a continuous series, there may be two methods of defining exact upper and lower limits of scores. Suppose a score or number is 61.

- Here, exact upper is 61.5 and lower limits is 60.5 of number or scores 61. It means 0.5 units below to 0.5 units above the face value of the score or number.
- Here, a score or number 61 means 61 not 62. A score or number of 61 represents any value between 61 and 62. Here, any fractional value less than 62 but greater than 61, will within the interval of 61-62, e.g. 61.2 or 61.6 or 61.8, is called 61.

Both the methods of defining exact upper and lower limits of scores are correct but it is safer to indicate clearly about the method to be used in proposed study.

Let us check our progress-2

- Differentiate between continuous and discrete series of data.*
- Describe level of measurement.*

5. Descriptive statistics

5.1 The frequency distribution

The researcher collects the data. These data are called raw data. There is a need of data to arrange them in a meaningful system, otherwise it has little meaning. Therefore, we arrange/ classify our data/ material into groups. This arrangement in a tabular or graphical format within a given class intervals are known as the frequency distribution.

**Table 6.3 The Original Scores of Obtained by
50 Students in College of Nursing in Statistics**

175 156 166 135 156 181 167 154 171 164 137 168 166 132 170 148 171 157 170 168
163 138 158 178 171 162 155 159 163 174 165 146 148 177 146 162 152 183 163 173
187 171 141 151 143 162 152 169 178 169

Table 6.4 The Grouped Scores into a Frequency Distribution

S.No.	Class Intervals	Tallies	Frequency (f)
(1)	(2)	(3)	(4)
1.	185-190	/	1
2.	180-185	//	2
3.	175-180	////	4
4.	170-175	///	5
5.	165-170	///	8
6.	160-165	///	10
7.	155-160	///	6
8.	150-155	////	4
9.	145-150	////	4
10.	140-145	//	2
11.	135-140	///	3
12.	130-135	/	1
			N=50

Note: In a frequency distribution table, when one digit is shown in two class intervals, then by the use of exclusive data method, the upper digit data will be placed in upper class interval.

The following are the rules for classifying scores to draw a frequency distribution:

(i) Determining the range or the gap between the lowest and the highest scores-

In Table 6.3, there is highest score is 187 and the lowest score is 132. The range can be calculated as $187-132=55$.

(ii) Fix the number and size of the groups for classification-

Commonly chosen grouping intervals are 3, 5, 10 units in length. If we divide the range (55) by 3, 5 and 10, it will result as 18.33, 11 and 5.5 respectively. An interval of 3 would spread (18.33=19 class interval) the data, we would not get the benefit of grouping. An interval of 10 would shrink (5.5= 6 class interval) the data, we would not get the benefit of grouping. If we see the class interval of 5, it seems suitable, so it is chosen in above Table 6.4.

(iii) The class intervals have been listed serially from the smallest scores at the bottom of the column to the largest scores at the top in column number (2) of the Table 6.4

In Table 6.4, first interval exists "130-135" and begins with score 130 and ends 134. It includes the 5 scores 130, 131, 132, 133 and 134. The topmost interval exists "185-190" and begins with score 185 and ends 190. It includes the 5 scores 185, 186, 187, 188 and 189. Tallies are marked in their respective interval. The rest scores are tabulated in the same manner. After writing of all 50, the total number of tallies are marked in each class interval in the column number (4), called as "f" (frequency). The sum of the "f" (frequency) is called N.

(iv) The limits of the class intervals in a frequency distribution

The following Table 6.5, describing the limits of the class intervals in a frequency distribution. In (Pattern-I), the class interval "130-135" explains that all scores from 130 up to 135 but it does not include 135 within this group. In (Pattern-II), the class interval "130-135", we see that all scores cover the same distance as described in (Pattern-I), but upper and lower limits of each interval are defined more exactly. In ((Pattern-III), the class interval shows the same facts more clearly than in (Pattern-I), but less exactly than in ((Pattern-II). When rapid tabulation of scores is needed, Pattern-III is suitable in comparison of method in (Pattern-I) and ((Pattern-II).

Table 6.5 The Limits of the Class Intervals into a Frequency Distribution

	Pattern-I			Pattern-II			Pattern-III		
S.No.	Class Intervals	Mid Point	(f)	Class Intervals	Mid Point	(f)	Class Intervals	Mid Point	(f)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1.	185-190	187	1	184.5-189.5	187	1	185-189	187	1
2.	180-185	182	2	179.5-184.5	182	2	180-184	182	2
3.	175-180	177	4	174.5-179.5	177	4	175-179	177	4
4.	170-175	172	5	169.5-174.5	172	5	170-174	172	5
5.	165-170	167	8	164.5-169.5	167	8	165-169	167	8
6.	160-165	162	10	159.5-164.5	162	10	160-164	162	10
7.	155-160	157	6	154.5-159.5	157	6	155-159	157	6
8.	150-155	152	4	149.5-154.5	152	4	150-154	152	4
9.	145-150	147	4	144.5-149.5	147	4	145-149	147	4
10.	140-145	142	2	139.5-144.5	142	2	140-144	142	2
11.	135-140	137	3	134.5-139.5	137	3	135-139	137	3
12.	130-135	132	1	129.5-134.5	132	1	130-134	132	1
			N=50			N= 50			N= 50

(v) The midpoint of the class intervals in a frequency distribution

In a frequency distribution, the scores grouped within an interval are considered to be spread over the entire class interval. If we want to represent all the scores, within a given class interval, by some single value, we have to calculate the midpoint of the class intervals in a given frequency distribution. This midpoint will represent all the scores within a given class interval. The formula of calculating of midpoint of interval is given below.

The midpoint of interval = Lower Limit of class interval + (Upper Limit - Lower Limit)/2
 For example, from our Table 6.5, row number 1, pattern-II, when exact limits are used in the frequency distribution.

Table 6.6 Pattern of Class Interval

	Pattern-I			Pattern-II			Pattern-III		
1.	185-190	187	1	184.5-189.5	187	1	185-189.5	187	1

$$\begin{aligned}
 \text{The midpoint of interval} &= 184.4 + (189.5 - 184.5)/2 \\
 &= 184.5 + 5/2 \\
 &= 184.5 + 2.5 = 187
 \end{aligned}$$

When we use the scores of the frequency distribution (pattern-III),

The midpoint of interval = beginning of class interval score + (upper Score - lower score)/2

$$\text{The midpoint of interval} = 185 + (189 - 185)/2$$

$$\begin{aligned} &= 185 + 4/2 \\ &= 185 + 2 = 187 \end{aligned}$$

The assumption of midpoint as representative of class interval in a frequency distribution, exists true, when large number of scores and small class intervals only.

5.2 Graphic presentation of the frequency distribution

The researcher translates numerical data or facts into more concrete and understandable forms, i.e. graphical method. These graphical methods may be as the frequency polygon, the histogram, the cumulative frequency graphs and cumulative curve or ogive.

(a) General characteristics

1. The graphs are made with reference to two lines.
2. These lines are known as coordinate axes.
3. The vertical line is called Y-axis; the second horizontal line is called X-axis.
4. These two lines (X-axis & Y-axis) are perpendicular to each other and they intersect at "O" or the origin point.
5. This system (intersection) forms four division or quadrants.
6. In upper right or first quadrant, measures positive on both axis (+ +).
7. In upper left or second quadrant, measures minus or negative on X-axis and positive on Y-axis (- +).
8. In lower left or third quadrant, measures minus or negative both axis (- -).
9. In lower right or fourth quadrant, measures positive on X-axis and on minus or negative Y-axis (- +).
10. The distance from "O" point on X-axis is known as abscissa and the distance from "O" point on Y-axis is known as ordinate.
11. It is easily understood by everyone without any prior knowledge.
12. It saves time.
13. It facilitates us to relate and compare the data for different time periods.
14. It is used in statistics to determine the mean, median and mode for different data.
15. It is also used in the interpolation and the extrapolation of data.

(b) The following points may be considered while making graphs.

1. There should be a suitable title of graph indicating the subject.
2. There should be measurement units in the graph.
3. There should be selected a proper scale.
4. There should be simple to understand everyone.
5. There should be correct size, fonts, colours, shades, lines, and design in the graphs.
6. The source of data should also indicate at the bottom of graph, (whenever it is necessary).

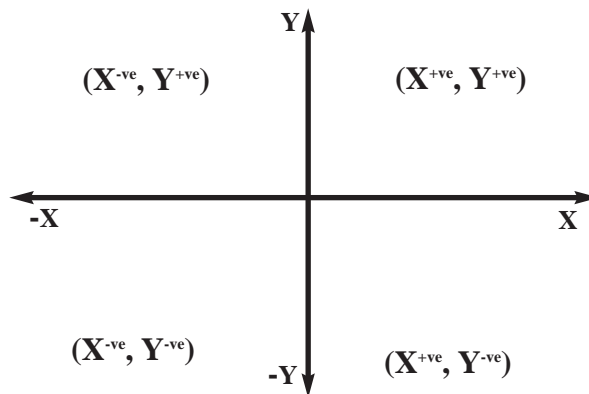


Figure : 6.1 Principle of graphical representation

(c) Various types of graphical representation

It is a way of analysing numerical data. It shows the relationship between data, ideas, information and concepts in a diagram. There are different ways of graphical representation. The main types of graphical representations are as follows:

(i) Line graphs - When the time series shows a wide range of fluctuations, the line graphs are used. It is used to display the continuous data. It is useful to predict future events over time.

(ii) Bar graphs - There are two types of bar graphs are used- the horizontal and the vertical bar graphs. It is used to display the category of data. It is used to compare the data. It shows solid bars to represent the quantities. The horizontal bars are used to show qualitative data where the vertical bars are used to show quantitative data or time series data.

(iii) Histograms - It uses bars to represent the frequency of numerical data that are organized into intervals. The

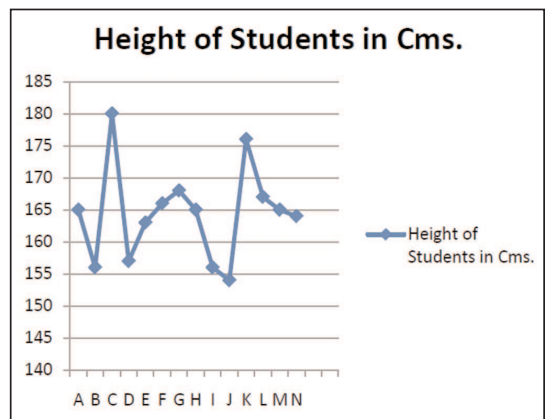


Figure : 6.2 Line graph

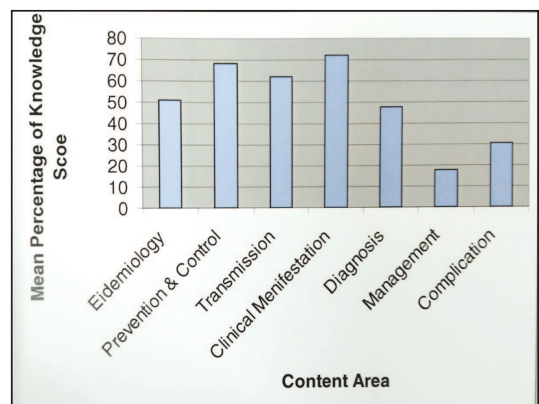


Figure : 6.3 Bar graph

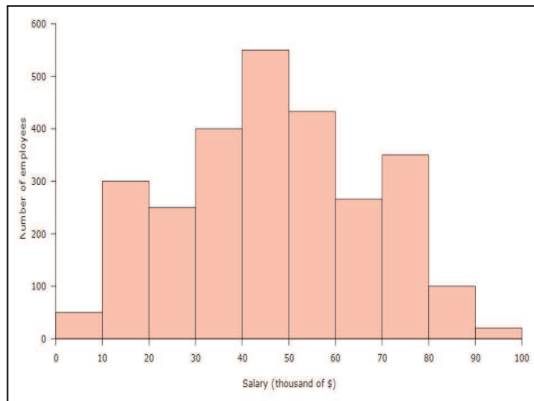


Figure : 6.4 Histograms graph

bars are placed continuously, side by side without any gap. Since all the intervals are equal and continuous, all the bars have the same width. The rectangles of histogram are erected on the class intervals of the distribution. The areas covered in rectangles are proportional to the frequencies.

(iv) Circle graph or pie chart - The whole circle represents the sum value of all components and the components or sectors of a circle represent their own values. The total angle of 360° the centre of circle is divided according to the value of the components/sectors. It shows the relationships of the parts of the whole.

The central angle of a component = $\frac{\text{Value of the component}}{\text{Total value}} \times 360$

When, the value of the components is expressed in percentages. The central angle of a component = $\frac{\text{Percentage value of the component}}{100} \times 360$.

The circle is considered with 100% and the categories occupied are represented with that specific percentage, e.g. 25%, 45%, 67%, etc.

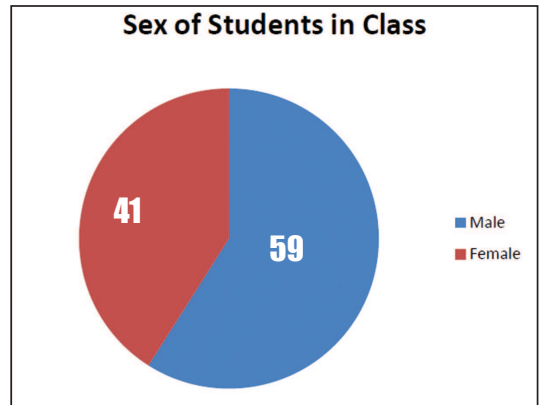


Figure : 6.5 Pie graph

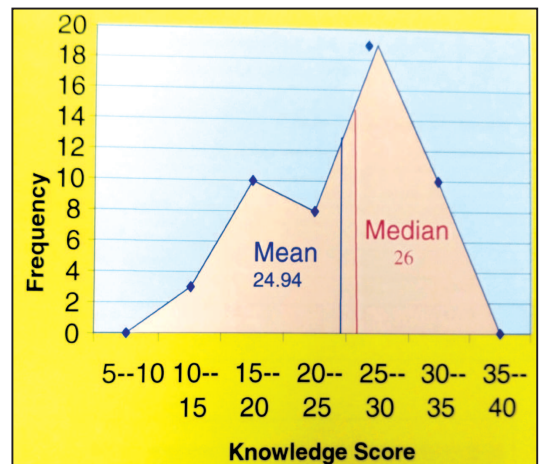


Figure : 6.6 Frequency polygon

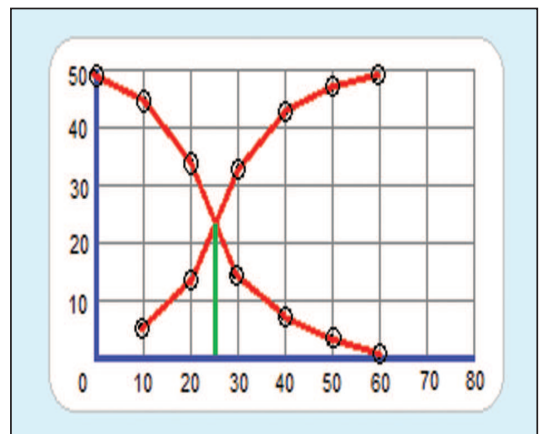


Figure : 6.7 Cumulative frequency graph

(v) The frequency polygon- It is also a method of graphical representation of frequency distribution. Firstly, the frequency distribution is prepared. The mid points of each class interval are calculated. The mid points on the X-axis and the frequencies on the Y-axis are placed. The placed points are joined by straight lines. Join each end point immediately to the lower or higher class marks on the X-axis. It is done to complete the polygon.

(vi) Ogives or cumulative frequency graphs- By plotting cumulative frequency against the respective class boundary, we get ogives. As such there are two ogives - less than type ogives, obtained by taking less than cumulative frequency on the vertical axis and more than type ogives by plotting more than type cumulative frequency on the vertical axis and there after joining the plotted points successively by line segments.

(d) Table 6.7 Difference Between the Frequency Polygons and Histograms

S. No.	Frequency polygons	Histograms
1.	It is a line graph.	It is a collection of adjacent rectangles.
2.	It is a multi-dimensional.	It is a two-dimensional.
3.	In this graph, several frequency distributions can be plotted on the same axis.	In this graph, we must have a separate graph for each distribution.
4.	It is a continuous curve and easy to determine the scope and change estimates.	It is not possible in a histogram.

5.3 Measures of central tendency

After research, we get some data or measures. These measures are tabulated into a frequency distribution, as shown in Table 6.3 and 6.4. Usually the next step is to calculate a central position or central tendency. It is an estimate of the centre of distribution values of a measure. It has two-fold-

- (i) It is an average of all the scores. It is a concise description of the performance by the group as a whole.
- (ii) It gives us ability to compare two or more groups. There are three types of averages or a central position or central tendency- the arithmetic mean, the median and the mode.

5.3.1 The Arithmetic Mean

The Arithmetic mean or mean or average is the most commonly used method of showing central tendency. To calculate the mean, all the values or scores are added and then divided by the number of values or scores.

(i) Calculation of the mean when data are ungrouped

Table 6.8 The Ungrouped Scores of Students in Statistics

175 156 166 135 156 181 167 154 171 164 137 168 166 132 170 148 171 157 170
168 163 138 158 178 171 162 155 159 163 174 165 146 148 177 146 162 152 183
163 173 187 171 141 151 143 162 152 169 178 169

Formula for the mean (M), when data are ungrouped

$M = \Sigma X / N$ (Here, M stands for mean, the symbol Σ stands for 'sum of', X stands for score or other measure and N is the number of measures in the series)

Now, we calculate the mean for the ungrouped scores given in Table 6.8

$M = \Sigma X / N$ (Here, $\Sigma X = 8071$ and $N = 50$)

$$= 8071 / 50$$

$$= 161.42$$

(ii) Calculation of the mean when data are grouped

Table 6.9 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Mid-Point(X)	(f)	(f X)
(1)	(2)	(3)	(4)	(5)
1.	185-190	187	1	187
2.	180-185	182	2	364
3.	175-180	177	4	708
4.	170-175	172	5	860
5.	165-170	167	8	1336
6.	160-165	162	10	1620
7.	155-160	157	6	942
8.	150-155	152	4	608
9.	145-150	147	4	588
10.	140-145	142	2	284
11.	135-140	137	3	411
12.	130-135	132	1	132
			N=50	8040

Formula for the mean (M), when data are grouped into a frequency distribution

$M = \Sigma fX / N$ (Here, M stands for mean, the symbol Σ stands for 'sum of', fX stands for the sum of the midpoints multiplied by their frequencies and N is the number of measures in the series)

Now, we calculate the mean for the grouped scores given in Table 6.9

First, we calculate midpoint (X) for each class interval. The scores grouped into class intervals lose their own identity, so they are represented by the particular interval on which they fall. The fX is calculated by multiplying the midpoint of each class interval by their respective frequencies. The fX is added and divided by N. We will obtain mean as a result of this process.

$M = \Sigma fX / N$ (Here, $\Sigma fX = 8040$ and $N = 50$)

$$= 8040 / 50$$

=160.8

(iii) Calculation of the mean when data are from combined grouped

Imagine that on the statistics subject. The test was conducted for two groups of students. The mean of first group of 15 students is 58 and the mean of second group of 12 students is 45. To calculate the combined mean, the following formula is used:

$$M \text{ (Combined)} = \frac{N_1M_1 + N_2M_2 + \dots + N_nM_n}{N_1 + N_2 + \dots + N_n}$$

(Here, M stands for mean, N is the number of measures in the series and n is number of groups)

In our imagination, there are two groups present, so formula may be as:

$$\begin{aligned} M \text{ (Combined)} &= (N_1M_1 + N_2M_2) / (N_1 + N_2) \text{ (Here, } N_1=15, M_1= 58, N_2 = 12, M_2 = 45) \\ &= (15 \times 58) + (12 \times 45) / (15 + 12) \\ &= (870 + 540) / 27 \\ &= 1410 / 27 \\ &= 52.22 \end{aligned}$$

Merits/uses of the mean:

- (i) It is easy to calculate.
- (ii) It is calculated from all values of distributions.
- (iii) It is an exact figure.
- (iv) It is least affected by fluctuations of sampling.
- (v) When the frequency distribution is not badly skewed or it is symmetrical, then, the mean is the centre of gravity in the frequency distribution.
- (vi) The mean is used, when the measure of central tendency having greatest stability is needed.
- (vii) When other statistics are not calculated yet. It is the base of many statistics.

Demerit of the mean:

- (i) It is affected by the extreme values of the distribution. It means, the smallest and the largest values affect the calculation of mean, e.g. the marks obtained by six nursing students in statistics are 30, 32, 33, 36, 38 and 98. The mean of this distribution is 44.5, which does not truly represent the group. The mean is largely affected by the highest value, 98. In another example, the marks obtained by six nursing students in statistics are 2, 40, 44, 46, 48 and 50. The mean of this distribution is 38.33, which does not truly represent the group. The mean is largely affected by the smallest value, 2.
- (ii) It cannot be estimated by looking the construction of distribution like median and mode.
- (iii) It is a fictitious figure; it may not be an actual item of the distribution.
- (iv) Sometimes it gives meaningless results, e.g. in our hospital, average 30.4 patients admitted in male medical ward in year 2021. Here, we can see that 30.4 patients can not be possible; the patients are living beings, which could

not be divided.

5.3.2 Median

The median is the midpoint or mid score in a series that is arranged in order of size. It is the score found at the exact middle of the set of values. It is not affected much by extreme items.

I. Calculation of the median from ungrouped data.

There may be two situations create in the calculation of the median from ungrouped data.

(i) When N is odd

Imagine that the students of B.Sc. nursing got the marks in a class test in the subject of statistics as given below:

6,8,9,12,11,14,7,8,10

To calculate the median from the above test scores, we have to arrange these nine scores in order of size: 6, 7, 8, 8, [9], 10, 11, 12, 14,

Here, midpoint is the 9, four scores are above 9 and four are below 9.

Hence, median is 9.

(ii) When N is even

Imagine that the students of B.Sc. nursing got the marks in a class test in the subject of statistics as given below:

8, 9, 12, 11,14,7,8,10

To calculate the median from the above test scores, we have to arrange these eight scores in order of size: 7, 8, 8, 9, 10, 11, 12, 14,

We use formula to find out median from ungrouped scores is given below:

Median (Mdn) = the $(N+1)/2^{\text{th}}$ measure in order of size

$$= 8+1/2$$

$$= 4.5^{\text{th}} \text{ score in size of order}$$

Here, the 4.5th score in size of order is 4th score (9) & 5th score (10). The midpoint is the 9.5 because it is higher limit of score 9 and lower limit of score 10.

Hence, median is 9.5.

II. Calculation of the median from grouped data.

When scores are available in a continuous series and grouped into a frequency distribution. The median will be 50% point in distribution. The following is the example of understanding the process of finding out the median.

Table 6.10 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Mid Point(X)	(f)	Counting 50% point	(f X)
(1)	(2)	(3)	(4)		(5)
1.	185-190	187	1		187
2.	180-185	182	2		364
3.	175-180	177	4		708
4.	170-175	172	5		860
5.	165-170	167	8		1336
6.	160-165	162	10		1620
7.	155-160	157	6	20	942
8.	150-155	152	4	14	608
9.	145-150	147	4	10	588
10.	140-145	142	2	6	284
11.	135-140	137	3	4	411
12.	130-135	132	1	1	132
			N=50		8040

In the above Table 6.10 to find out the median it is taken 50% of our scores ($N/2=25$). We begin counting from lower side or small score. We start at the class interval 130-135 and adding up scores serially, we reach the class interval 155-160. Here, we find 20 frequencies, it is short to 25 frequencies, which are necessary to find out the median ($N/2=25$). The next class interval is 160-165 which contain 10 frequencies. It is assumed that the median will present in this class interval. To fulfill the requirement of five frequencies, we take $5/10 \times 5$ (5=needed frequencies, 10=frequencies of next class interval and 5=length of the interval). By this calculation, we get 2.5 and add it in the 159.5, the beginning of the class interval 160-165. We get the median at $159.5 + 2.5=162$.

Formula for calculating the median (Mdn), when data are grouped into frequency distribution

$$Mdn = l + \left(\frac{\frac{N}{2} - F}{f_m} \right) \times i$$

(Here, Mdn stands for median, the l = exact lower limit of the class interval upon which the median lies, $N/2$ = one half of the total number of scores, F = sum of the scores on all intervals below l ,

f_m =frequency (number of scores) within the interval upon which the median fall, i =length of the class interval)

We will calculate the median from the data given in the Table 6.10

(Here, the l = 159.5, $N/2$ = $50/2=25$, F = 20, f_m = 10, i = 5)

$$\begin{aligned} Mdn &= 159.5 + (50/2 - 20/10) \times 5 \\ &= 159.5 + (25 - 20/10) \times 5 \end{aligned}$$

$$\begin{aligned}
 &= 159.5 + (5/10) \times 5 \\
 &= 159.5 + (0.5) \times 5 \\
 &= 159.5 + 2.5
 \end{aligned}$$

$$\text{Mdn} = 162$$

Merits/uses of the median:

- (i) It is easy to calculate and to understand.
- (ii) It is an actual figure of series.
- (iii) It is also appropriate measure in qualitative phenomena.
- (iv) It is always rigid and clear.
- (v) It is used when the exact midpoint (the 50% point) of frequency distribution is needed.
- (vi) In a frequency distribution, if there are extreme scores present, definitely these extreme scores would markedly affect the mean but not to the median. So, the median is preferred.
- (vii) When it is thought that certain scores should influence the central tendency. In such situation median is preferred.

5.3.3 Mode

(i) In ungrouped set of scores

Mode is the most frequently occurring value in any set of scores. To determine the mode, the scores are arranged in an order and then counted each one. The most frequently occurring value is the mode.

For example, 13 13 13 15 15 18 19 20 is a set of scores.

In this set of scores, only the score 13, occurs three times so it is the mode of this set of scores.

In some distributions there may be more than one mode values.

For example, 13 13 13 15 15 15 18 19 20 is a set of scores.

In this set of scores, the score 13 and 15 both, occurs three times so there is present of two modes. This condition is called, bimodal distribution.

For example, 13 13 13 15 15 15 18 19 19 19 20 is a set of scores.

In this set of scores, the score 13, 15 and 19, occurs three times so there is present of three modes. This condition is called, multimodal distribution.

(ii) In grouped set of scores

When data are grouped into the frequency distribution, we can take the crude or empirical mode from mid-point of that interval, which contains the largest frequency. We can refer Table 6.11, in this table; the class interval 160-165 contains the largest frequency (10). Its mid-point is (162 or 161.5), hence 161.5 is the crude mode of this grouped frequency distribution.

Table 6.11 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Mid Point(X)	Frequencies (f)
(1)	(2)	(3)	(4)
1.	185-190	187	1
2.	180-185	182	2
3.	175-180	177	4
4.	170-175	172	5
5.	165-170	167	8
6.	160-165	162	10
7.	155-160	157	6
8.	150-155	152	4
9.	145-150	147	4
10.	140-145	142	2
11.	135-140	137	3
12.	130-135	132	1
			N=50

After calculating the crude mode, now we will calculate the true mode. The true mode is the point of greatest concentration in the frequency distribution. It is calculated by the following formula:

Here,

l = Lower limit of the modal class interval (having maximum frequency)

f_m = Frequency of modal class interval

f_1 = Frequency of class interval preceding the modal class interval

f_2 = Frequency of class interval following the modal class interval

i = Width of the modal class interval

$$Mode = l + \left(\frac{f_m - f_1}{2f_m - f_1 - f_2} \right) \times i$$

Now, we will practice of the formula with the table 6.11. In this table, the maximum frequency is 10, which lies in class interval 159.5-164.5.

Here, l = Lower limit of the modal class interval (having maximum frequency) = 159.5

f_m = Frequency of modal class interval = 10

f_1 = Frequency of class interval preceding the modal class interval = 8

f_2 = Frequency of class interval following the modal class interval = 6

i = Width of the modal class interval = 5

Let us put in formula

$$\begin{aligned}
 \text{Mode} = l + \left(\frac{fm - f_1}{2fm - f_1 - f_2} \right) \times i &\Rightarrow \\
 &= 159.5 + \{ (10-8)/2 \times (10-8-6) \} \times 5 \\
 &= 159.5 + \{ 2/(20-8-6) \} \times 5 \\
 &= 159.5 + 2/6 \times 5 \\
 &= 159.5 + 10/6 \\
 &= 159.5 + 1.67 \\
 \text{Mode} &= 161.17
 \end{aligned}$$

When the frequency distribution is symmetrical, the formula is used to determine the mode.

Mode=3 Median-2 Mean

Merits/ uses of the mode:

- (i) It is easy to calculate and easy to understand.
- (ii) It is not affected by the extreme items of distribution.
- (iii) It is not affected by the dispersion of the series like median.
- (iv) When a quick and approximate score of central tendency is needed, the mode is the first preference.
- (v) When a measure of central tendency should need as a typical value, mode is taken.
- (vi) It is considered more reliable in commercial estimation and forecasting.
- (vii) It is used in weather forecasting for rainfall, temperature and wind direction.

5.4 Measures of variability

The central tendency are important measures and useful in describing the nature of a frequency distribution. They are unable to give a complete scene of data to the researcher. These measures will not show how the scores tend to be distributed. To solve this problem, we use measures of variability or spread or dispersion. The ranges, the variance and standard deviation are most commonly used measures of variability.

5.4.1 The Range

The range is the difference between the two extreme measures or values in a distribution. It is interval between highest score and the lowest score. It is the most general measure of scatter or spread or dispersion. We calculate range, when we need a rough comparison of two or more groups. Imagine the following are the scores of B.Sc. nursing students in statistics.

48, 38, 37, 27, 26, 28, 19, 18.

We calculate the range of the above scores. We will find out the difference between 48 and 18, it is $48-18=30$

Here, the range is easily calculated but it has some important limitations, as follows:

- (a) The range is the result of two extreme values in the distribution. It does not show the variation of other scores in a distribution.
- (b) It does not give stable statistics; it can differ from sample to sample, even in the same population.
- (c) It is unreliable when N is small or has a large gap in the frequency distribution.

5.4.2 The standard deviation or SD

The concept of standard deviation was first introduced in 1893 by Karl Pearson. The standard deviation or SD is a more accurate, important, popular and detailed estimate or meas-

ure of dispersion or variability. It is the most stable index of dispersion or variability. It is used in experimental and research work. The squared deviations are used in calculating the SD. The standard deviation or SD is the positive squared root of variance. These are taken from the mean only not from median or mode. It is represented by the Greek letter sigma (σ).

(i) (a) Calculation of the standard deviation from the ungrouped set of scores:

With the help of deviation scores

- We take the following set of ungrouped set of data:
16,18,19,15, 21, 17, 22, 16
- Calculate the arithmetic mean of above set of scores ($M=\Sigma X/N=144/8=18$)
- Find the value of x, the mean (18) is subtracted from raw scores. The values that are below the mean have negative deviations/ discrepancies and values above the mean are positive.
- Square each deviation
- Sum these squares and get sum of squares value. ($\Sigma x^2 = 44$)

Table 6.12 Computing Standard Deviation from the Ungrouped set of Scores

S.No.	Scores(X)	Scores-mean (X-M=x) (Deviation from the mean) (x)	Squared of deviation(x ²)
1.	16	16-18= -2	4
2.	18	18-18= 0	0
3.	19	19-18= 1	1
4.	15	15-18= -3	9
5.	21	21-18= 3	9
6.	17	17-18= -1	1
7.	22	22-18= 4	16
8.	16	16-18= -2	4
			$\Sigma x^2 = 44$

- Apply the formula for calculation of standard deviation (To get the standard deviation, we take the square root of the variance because we have squared the deviations earlier).

$$\begin{aligned}
 \text{(Here, } \Sigma x^2 = 44 \text{ and } n=8) \quad \sigma &= \sqrt{\frac{\Sigma X^2}{(n-1)}} = \sqrt{\frac{44}{(8-1)}} \\
 &= \sqrt{6.2857} \Rightarrow 2.5071
 \end{aligned}$$

If the sample size (n) is small ($n < 30$) then sample standard deviation is obtained by dividing the sum of squares of the deviations by (n-1) instead of (n).

(i) (b) Calculation of the variations and standard deviation from the ungrouped set of scores: With the help of raw scores

Table 6.13 Computing SD from the Ungrouped set of Scores

S.No.	Scores (X)	Squared of Scores (X ²)
1.	16	256
2.	18	324
3.	19	361
4.	15	225
5.	21	441
6.	17	289
7.	22	484
8.	16	256
	$\Sigma X=144$	$\Sigma X^2 = 2636$

The formula used for calculating standard deviation for ungrouped set of data with the help of raw data is: Standard Deviation= (σ) = $\sqrt{\{N \cdot \Sigma X^2 - (\Sigma X)^2\} / N}$
(X= Raw measures/scores, N= the number of scores in the distribution)

a. We take the following set of ungrouped set of data:

16, 18, 19, 15, 21, 17, 22, 16

b. Calculate the sum of scores is = $\Sigma X=144$

c. Calculate the sum of squared scores is = $\Sigma X^2 = 2636$

d. Apply the formula for calculation of standard deviation

$$\text{Standard Deviation} = (\sigma) = \sqrt{\{N \cdot \Sigma X^2 - (\Sigma X)^2\} / (n-1)}$$

$$= \sqrt{8 \times 2636 - (144)^2 / (n-1)}$$

$$= \sqrt{21088 - 20736 / (8-1)}$$

$$= \sqrt{352 / 7}$$

$$= 18.761663 / 7$$

$$= 2.6802$$

Here, $n < 30$, so denominator will $n-1$, instead of n

(ii) (a) Calculation of the standard deviation from the grouped set of scores: With the help of assumed mean deviation scores

Table 6.14 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Mid-Point (x)	(f)	(x')	(fx')	(fx' ²)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.	185-190	187	2	5	10	50
2.	180-185	182	2	4	8	32
3.	175-180	177	3	3	9	27
4.	170-175	172	5	2	10	20
5.	165-170	167	8	1	8 (+45)	8

contd. on next page

S.No.	Class Intervals	Mid-Point (x)	(f)	(x')	(fx')	(fx' ²)
6.	160-165	162	9	0	0	0
7.	155-160	157	6	-1	-6	6
8.	150-155	152	4	-2	-8	16
9.	145-150	147	5	-3	-15	45
10.	140-145	142	2	-4	-8	32
11.	135-140	137	2	-5	-10	50
12.	130-135	132	2	-6	-12	72
			N=50		Σfx'=-14	Σfx'²=358

The formula used for calculating variance for grouped set of data with the help of assumed mean deviation scores is: Standard Deviation= $(\sigma) = i / N [\sqrt{N \Sigma fx'^2} - (\Sigma fx')^2]$ (i= Width of the class-interval, x'= Deviation of the raw measures/scores from the assumed mean,

N= The number of scores in the distribution)

a. We take the grouped set of data in Table 6.14:

b. Calculate deviation of raw data from the assumed mean

c. Calculate the sum of deviation of raw data from the assumed mean is =fx'

d. Calculate the sum of squared deviation of raw data from the assumed mean = fx'²

e. Apply the formula for calculation of standard deviation

Standard Deviation= $(\sigma) = i / N [\sqrt{N \Sigma fx'^2} - (\Sigma fx')^2]$

(i= Width of the class-interval=5, x'= Deviation of the raw measures/scores from the assumed mean = -23, sum of squared deviation of raw data from the assumed mean =363, N= The number of scores in the distribution=50)

Standard Deviation= $(\sigma) = i / N [\sqrt{N \Sigma fx'^2} - (\Sigma fx')^2]$

$$= 5/50 [\sqrt{50 \times 358 - (-14)^2}]$$

$$= 5/50 [\sqrt{17900 - 196}]$$

$$= 5/50 \sqrt{17704}$$

$$= 1/10 \times 133.0563$$

Standard Deviation= $(\sigma) = 13.3056$

Let us check our progress-3

1. Calculate the mean for the following frequency distribution.

(a) 75 56 66 35 56 81 67 54 71 64 37 68 66 32 70 48 71 57 70 68 63 38 58 78 71

(b) 62 55 59 63 74 65 46 48 77 46 62 52 83 63 73 87 71 41 51 43 62 52 69 78 69

(c) 17 15 16 13 15 18 16 15 17 16 13 16 16 13 17 14 17 15 17 16 16 13 15 17 17

2. Calculate the mean for the following frequency distribution.

(a) Table 6.15 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Frequencies
1.	180-185	1
2.	175-180	2
3.	170-175	4
4.	165-170	5
5.	160-165	8
6.	155-160	10
7.	150-155	6
8.	145-150	4
9.	140-145	4
10.	135-140	2
11.	130-135	3
12.	125-130	1

(b) Table 6.16 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Frequencies
1.	175-180	1
2.	170-175	2
3.	165-170	4
4.	160-165	5
5.	155-160	8
6.	150-155	10
7.	145-150	6
8.	140-145	4
9.	135-140	4
10.	130-135	2
11.	125-130	3
12.	120-125	1

(c) Table 6.17 The Data Grouped into a Frequency Distribution

S. No.	Class Intervals	Frequencies
(1)	(2)	(4)
1.	170-175	1
2.	165-170	2
3.	160-165	4
4.	155-160	5
5.	150-155	8
6.	145-150	10
7.	140-145	6
8.	135-140	4
9.	130-135	4
10.	125-130	2
11.	120-125	3
12.	115-120	1

3. Calculate the median for the following frequency distribution.

(a) 75 56 66 35 56 81 67 54 71 64 37 68 66 32 70 48 71 57 70 68 63 38 58 78 71

(b) 62 55 59 63 74 65 46 48 77 46 62 52 83 63 73 87 71 41 51 43 62 52 69 78 69

(c) 17 15 16 13 15 18 16 15 17 16 13 16 16 13 17 14 17 15 17 16 16 13 15 17 17

4. Calculate the median for the following frequency distribution.

(a) Table 6.18 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Frequencies (f)
(1)	(2)	(3)
1.	180-185	1
2.	175-180	2
3.	170-175	4
4.	165-170	5
5.	160-165	8
6.	155-160	10
7.	150-155	6
8.	145-150	4
9.	140-145	4
10.	135-140	2
11.	130-135	3
12.	125-130	1

(b)Table 6.19 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Frequencies (f)
(1)	(2)	(4)
1.	175-180	1
2.	170-175	2
3.	165-170	4
4.	160-165	5
5.	155-160	8
6.	150-155	10
7.	145-150	6
8.	140-145	4
9.	135-140	4
10.	130-135	2
11.	125-130	3
12.	120-125	1

(c) Table 6.20 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Frequencies(f)
(1)	(2)	(4)
1.	170-175	1
2.	165-170	2
3.	160-165	4
4.	155-160	5
5.	150-155	8
6.	145-150	10
7.	140-145	6
8.	135-140	4
9.	130-135	4
10.	125-130	2
11.	120-125	3
12.	115-120	1

5. Calculate the mode for the following frequency distribution.

(a) 75 56 66 35 56 81 67 54 71 64 37 68 66 32 70 48 71 57 70 68 63 38 58 78 71

(b) 62 55 59 63 74 65 46 48 77 46 62 52 83 63 73 87 71 41 51 43 62 52 69 78 69

(c) 17 1516 13 15 18 16 15 17 16 13 16 16 13 17 14 17 15 17 16 16 13 15 17 17

6. Calculate the mode for the following frequency distribution.

(a) Table 6.21 The data grouped into a frequency distribution

S.No.	Class Intervals	Frequencies (f)
(1)	(2)	(3)
1.	180-185	1
2.	175-180	2
3.	170-175	4
4.	165-170	5
5.	160-165	8
6.	155-160	10
7.	150-155	6
8.	145-150	4
9.	140-145	4
10.	135-140	2
11.	130-135	3
12.	125-130	1

(b) Table 6.22 The data grouped into a frequency distribution

S.No.	Class Intervals	Frequencies (f)
(1)	(2)	(3)
1.	175-180	1
2.	170-175	2
3.	165-170	4
4.	160-165	5
5.	155-160	8

Contd...

6.	150-155	10
7.	145-150	6
8.	140-145	4
9.	135-140	4
10.	130-135	2
11.	125-130	3
12.	120-125	1

(c) Table 6.23 The data grouped into a frequency distribution

S.No.	Class Intervals	Frequencies (f)
(1)	(2)	(3)
1.	170-175	1
2.	165-170	2
3.	160-165	4
4.	155-160	5
5.	150-155	8
6.	145-150	10
7.	140-145	6
8.	135-140	4
9.	130-135	4
10.	125-130	2
11.	120-125	3
12.	115-120	1

7. Calculate the standard deviation from the following ungrouped data.(Take the help of deviation scores)

(a) 26 28 29 25 24 27 23 26

(a) 20 25 28 25 24 26 24 26 29 23

(b) 16 18 19 15 14 17 13 16

8. Calculate the standard deviation from the following ungrouped data.(Take the help

of raw scores)

(a) 26 28 29 25 21 27 22 26

(b) 16 18 19 15 11 17 12 16 18

(c) 16 18 19 15 11 17 12 16 18 17

9. Calculate the standard deviation from the following grouped set of scores.(Take the help of assumed mean)

(a) Table 6.24 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Frequencies (f)
(1)	(2)	(3)
1.	175-180	1
2.	170-175	2
3.	165-170	4
4.	160-165	5
5.	155-160	8
6.	150-155	10
7.	145-150	5
8.	140-145	4
9.	135-140	4
10.	130-135	2
11.	125-130	3
12.	120-125	2

(b) Table 6.25 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Frequencies (f)
1.	170-175	1
2.	165-170	1
3.	160-165	2
4.	155-160	3
5.	150-155	4
6.	145-150	7
7.	140-145	2
8.	135-140	2
9.	130-135	2
10.	125-130	3
11.	120-125	2
12.	115-120	1

(c) Table 6.26 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Frequencies (f)
(1)	(2)	(3)
1.	165-170	2
2.	160-165	2
3.	155-160	2
4.	150-155	3
5.	145-150	3
6.	140-145	6
7.	135-140	2
8.	130-135	1
9.	125-130	2
10.	120-125	3
11.	115-120	2
12.	110-115	2

6. Normal probability

Normal Probability distribution specifies how the values of a variable are distributed. It is a symmetric, most of the observations gather around the central peak of curve. It is based on the law of probability. It is a mathematical model, not an actual distribution of measures or scores. It is represented by a curve. This curve is

known as "Normal Probability Curve." It is also called Gaussian distribution. This curve has shape of bell, so sometimes it is called a bell curve. A large number of random variables are either nearly or exactly represented by the normal distribution.

The Properties of Normal Distribution

1. In a normal distribution, all central tendencies (mean, median and mode) are equal.
2. Total area under the curve should be equal to 1.

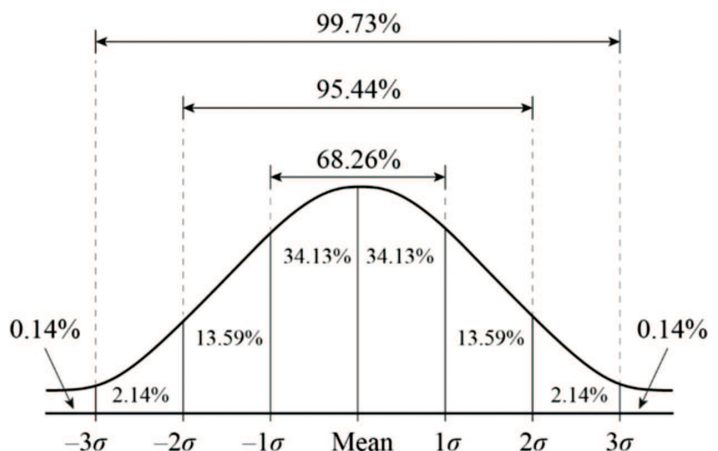


Figure 6.2 : Normal Probability Curve

3. It should be symmetrical around its ordinate (vertical axis). It means the size, shape and slope of the curve on one side of the ordinate should be identical to that on its other side or we can say that the exactly half of the values are to be the right of the centre and exactly half of the values are to be the left of the centre.
4. The height of the ordinate (vertical axis) is maximum at the mean. In the unit, the normal curve is equal to 0.3989.
5. It should have only one peak is called unimodal.
6. In a normal distribution, the curve is asymptotic. It approaches the x-axis. But it never meets horizontal axis. It extends from $-\infty$ (minus infinity) to $+\infty$ (plus infinity).
7. The point of inflection of the curve occurs at point's ± 1 , standard deviation (± 1 sigma), above and below the mean. Thus, the curve changes from convex to concave in respect to the horizontal axis at these points.
8. About 68.26 % of the total area falls between the limits $M \pm 1$ sigma.
9. About 95.44 % of the total area falls between the limits $M \pm 2$ sigma.
10. About 99.73 % of the total area falls between the limits $M \pm 3$ sigma.

7. Test of significance

7.1 Meaning

"Tests of significance" is a very common. It is an important term in a statistical process. The statistical significance is the probability of a result of some statistical test. The researcher has to make sure about the quality of sample, accuracy and good measures. The researcher has to determine if the findings of experiments have occurred as a result of a study or just by chance. Tests of significance are the method of rejecting or accepting the claims based on sample data. There are many methods to test the significance, known as test of significance. This testing may have some degree of error. In some experiments, the researcher has to define the probability of sampling error in advance. In any study, where does not include the entire population, the sampling error does exist.

Significance tests tell us about the probability. It says that if a relationship is due to random chance or not. It also denotes the level of relationship. It indicates about the error that would be made by us if they found relationship is assumed to exist.

7.2 p-values

The p-value is a function of observed sample results. It is being used for testing of statistical hypothesis. It is to be selected before the test is performed. Traditionally, it is kept 1% or 5% and this value is known as the significance level. It is denoted by α (alpha).

When the p-value is smaller than or equal to significance level α (alpha), the data is said to be inconsistent for our assumption of the null hypothesis to be true.

Therefore, the null hypothesis should be rejected and an alternative hypothesis is supposed to be accepted or assumed as true.

Note that the smaller the p-value is, the bigger the significance should be as it indicates that the research hypothesis does not adequately explain the observation. If the p-value is calculated accurately, then such test controls type I error rate not to be greater than the significance level α (alpha).

7.3 The significance levels

The significance level α (alpha) is the probability of making the wrong decision when the null hypothesis is true, in a hypothesis test. It is the level at which it can be accepted if a given event is statistically significant. This is also called p-value. The bigger samples are less prone to chance. The sample size plays an important role in measuring the statistical significance. Only representative and random samples should use for significance testing. When a researcher looks for a 5% probability or less which means 5% results occurs due to chance and there is a 95% chance of results occurring not by chance, it indicates that the result of our experiment is statistically significant.

7.4 The level of confidence

Confidence level is the probability. It says that if a test or survey is repeated over and over again, the results obtained would be the same. It is an expression of a researcher's confidence in obtaining data from a sample. It is expressed as a percentage. It indicates the percentage of the target population would give an answer that lies within the confidence interval.

7.5 Types of errors

There are basically two types of errors

(i) Type I

When the researcher finds that the assumed relationship through research hypothesis does exist; but in reality, it does not exist as evidence. In this type of error, the researcher is supposed to reject the research hypothesis and accept the null hypothesis, but its opposite happens. The probability that researchers commit Type I error. It is denoted by alpha (α).

(ii) Type II

It is just opposite the type I error. When it is assumed that a relationship does not exist, but in reality, it does. In this type of error, the researcher is supposed to accept the research hypothesis and reject the null hypothesis, but he does not and the opposite happens. The probability that a type II error is committed is represented by beta (β).

8. Co-efficient of correlation

8.1 Meaning

When we study related to one variable, it is called uni-variate distributions. When we study related to two variables, it is called bi-variate distributions. When we

study related to more than two variables, it is called multi-variate distributions. When the values of two variables vary in such a way that fluctuations in one are accompanied by fluctuations in the other, these two variables are said to be correlated. It is observed that better nursing care decreases the hospital stay of a patient and using of preventive measures of Covid-19 decreases the risk of the Covid-19 disease. Thus, it is said that two variables are inter-dependent on each other. Two variables are said to be correlated if the change in one variable result in a corresponding change in the other variable.

8.2 Types and degree of co-relation

Co-efficient of correlation is a measure of the degree of relationship between two sets of measures for the same group of individuals. This degree of relationship is known as correlation and it can be quantitatively represented by the co-efficient of correlation. Its value ranges from -1.00 to +1.00. A value of -1.00, denotes a complete negative relationship and to +1.00 indicating perfect positive relationship. A zero value (0.00) describes complete lack of relationship between the two variables. There are various methods of correlation of variables. The use of correlation is done according to situation and type of data.

Table 6.27 Value of Correlation

S.No.	Value	Denotes or Indicates
1.	0.00 to ± 0.20	Indifferent or negligible relationship
2.	± 0.20 to ± 0.40	Low correlation; present but slight
3.	± 0.40 to ± 0.70	Substantial or marked relationship
4.	0.70 to ± 1.00	High to very high relationship

8.3 Methods of determining co-relation

(A) Graphic methods: It is a method of determining co-relation, in which graph device is used to find out correlation between two variables.

(a) Scatters or dot diagram method: In this method, the values of independent data series (predicting variables) are plotted on O-X (horizontal) axis and those of dependent series (predicted variables) on O-Y (vertical) axis and then the pairs of values are plotted on the graph paper. As a result, graphs of dots are obtained. These dots are scattered in different forms. Therefore, the graphs are called scatter diagrams or dotograms. The patterns of scatter diagrams indicate the following views:

- ❖ If the dots of the two series are forwarding in a definite direction like a current, this condition indicates that the data series are definitely correlated.
- ❖ When the arrays of dots are facing toward advance from left to right in upward direction, the correlation is definitely positive.
- ❖ When the scatter diagram directing from left to right in downward direction,

the correlation is negative.

❖ When the dots are not in definite arrays and are scattered haphazardly, this condition indicates that there is no correlation between the data series.

❖ When the dots appear to be situated on a line which advances upward at 45° angle from the O-X axis, this condition indicates perfect positive correlation among the data series.

❖ If the dots appear to be situated on a line which moves from left to right in downward direction at 45° angle from O-X axis, this condition is indicative of perfect negative correlation.

(b) Simple graphic method: In this method, the values of dependent series are plotted on O-X axis and independent series are plotted on O-Y axis of graph paper, a linear or non-linear graph will be obtained which will simply indicate the direction of correlation and not the numerical magnitude.

❖ If the graph lines of two independent series move in upward direction from left to right, the correlation is called positive.

❖ If the graph line of one series moves upward from left to right and that of the other independent series moves downward from left to right, it is called negative correlations.

❖ If the values of two data series do not show either positive or negative trend then it should be inferred that there is no correlation.

(B) Mathematical methods

(a) Karl Pearson's product moment method: Coefficient of correlation (r) is obtained by dividing the product of values of covariance of the two series by the product of their standard deviations. When data for two variables (X and Y) are expressed in ratio or interval of measurement, distributions of variables are uni-modal; their variances are approximately equal and have linear relationship. In such situations this method of correlation is used. It is called Pearson's r.

(i) Calculation of Pearson's r from ungrouped data

When size of sample is small and data are not grouped then Pearson's r may be calculated as follows:

$$r_{xy} = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}}$$

Here, x = deviations of X measures from assumed mean

y = deviations of Y measures from assumed mean

We may take an example and calculate the Pearson's product moment 'r' from the following data of the two variables for 10 nursing students of our college.

Table 6.28 Two Variables for 10 Nursing Students of our College

X	55	64	62	69	72	56	65	59	60	64
Y	52	60	65	56	69	51	56	58	55	58

Calculation:

X	Y	x= X- AM	y = Y- AM	x ²	y ²	Xy
55	52	55-62= -7	52-58= -6	49	36	42
64	60	64-62= 2	60-58= 2	4	4	4
62 (AM)	65	62-62= 0	65-58= 7	0	49	0
69	56	69-62= 7	56-58= -2	49	4	-14
72	69	72-62= 10	69-58= 11	100	121	110
56	51	56-62= -6	51-58= -7	36	49	42
65	56	65-62= 3	56-58= -2	9	4	-6
59	58(AM)	59-62= -3	58-58= 0	9	0	0
60	55	60-62= -2	55-58= -3	4	9	6
64	58	64-62= 2	58-58= 0	4	0	0
		Σx =6	Σy =0	Σx ² =264	Σy ² =276	Σxy=184

$$\begin{aligned}
 r_{xy} &= \frac{N \cdot \Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[N \cdot \Sigma x^2 - (\Sigma x)^2][N \cdot \Sigma y^2 - (\Sigma y)^2]}} \\
 &= \frac{10 \times 184 - 6 \times 0}{\sqrt{[(10 \times 264) - (6)^2][10 \times 276 - (0)^2]}} \\
 &= \frac{1840 - 0}{\sqrt{[2640 - 36][2760 - 0]}} \\
 &= \frac{1840}{\sqrt{[2604][2760]}} \\
 &= \frac{1840}{\sqrt{7187040}} \\
 &= \frac{1840}{2680.8655} \\
 &= 0.6863
 \end{aligned}$$

(ii) Calculation of Pearson's r from grouped data

When the number of items are large and the data are grouped. These data are classified into two-way frequency distribution or a frequency table. It is also known as correlation table. It is a tabular method of finding correlation. It is applicable for two variables only. It is indicated for both discrete as well as continuous data. In table, one variable is shown in the rows and the other shown in the columns. The frequencies in each class are indicated in squares in body of the table. The total frequencies of each class are indicated in marginal rows and columns. The inner squares show the distribution of the total of each row into various columns and the total of each column into various rows. The example given below:

Table 6.29 Marks Obtained and Age of Students

Marks obtained in mid-term test	Age of students in years					Total
	18	19	20	21	22	
5-10	4	5	5	4	5	23
10-15	5	5	4	5	4	23
15-20	5	6	5	5	5	26
20-25	5	4	6	7	6	28
Total	19	20	20	21	20	100

Following steps may be taken for calculation of Pearson's r from grouped data:

- (i) First of all, we take the step deviations of x variable and denote these deviations by 'dx'.
- (ii) After it, we take the step deviations of y variable and denote these deviations by 'dy'.
- (iii) Multiply dx, dy and the respective frequency of each cell and write the figure obtained in the left-hand corner of each cell.
- (iv) Now we add together all the values as calculated in step
- (iii) and obtain the total $\Sigma f dx dy$.
- (v) Multiply the frequencies of x variable by dx and obtain the total $\Sigma f dx$.
- (vi) Multiply $f dx$ by dx and obtain the total $\Sigma f dx^2$.
- (vii) Multiply the frequencies of y variable by dy and obtain the total $\Sigma f dy$.
- (viii) Multiply $f dy$ by dy and obtain the total $\Sigma f dy^2$.
- (ix) Substitute the values of $\Sigma f dx dy$, $\Sigma f dx$, $\Sigma f dx^2$, $\Sigma f dy$, $\Sigma f dy^2$ and N in the following formula.

Formula: $r = \frac{\Sigma f dx dy \cdot N - (\Sigma f dx \cdot \Sigma f dy)}{\sqrt{[\Sigma f dx^2 \cdot N - (\Sigma f dx)^2][\Sigma f dy^2 \cdot N - (\Sigma f dy)^2]}}$

- (x) As a result, we obtain the value of 'r'.

Example: Calculate coefficient of correlation between marks obtained in Anatomy and Physiology of nursing students of a college of nursing in mid-term examination.

Table 6.30 Marks Obtained in Anatomy and Physiology

Marks obtained in Anatomy (Y)	Marks obtained in Physiology (X)					Total
	25-35	35-45	45-55	55-65	65-75	
20-30	6	3	4	-	-	13
30-40	3	16	10	5	-	34
40-50	-	10	14	7	-	31
50-60	-	-	8	9	3	20
60-70	-	4	-	4	5	13
Total	9	33	36	25	8	111

Marks obtained in Anatomy (Y)	Marks obtained in Physiology (X)									
	25-35	35-45	45-55	55-65	65-75	f	fdy	fd ² y	fdxdy	
	dx	-2	-1	0	+1	+2				
	dy									
20-30	-2	4	2	0						
		6	3	4	-	13	-26	52	30	
		24	6	0						
30-40	-1	2	1	0	-1					
		3	16	10	5	34	-34	34	17	
		6	16	0	-5					
40-50	0		0	0						
		-	10	14	7	31	0	0	0	
			0	0	0					
50-60	+1			0	1					
		-	-	8	9	20	20	20	15	
				0	9	6				
60-70	+2		-2	0	2					
		-	4	-	4	13	26	52	20	
			-8		8	20				
f		9	33	36	25	111(N)	(-)14Σfdy	158Σfd ² y	82Σfdxdy	
fdx		-18	-33	0	25	(-)10Σfdx				
fd ² x		36	33	0	25	126 Σfd ² x				
fdxdy		30	14	0	12	82Σfdxdy				

$$\begin{aligned}
r &= \frac{\sum fdxdy.N - (\sum fdx.\sum fdy)}{\sqrt{\sum fd^2x.N - (\sum fdx)^2} \sqrt{\sum fd^2y.N - (\sum fdy)^2}} \\
&= \frac{(82 \times 111) - \{(-10) \times (-14)\}}{\sqrt{[126 \times 111 - (-10)^2]} \sqrt{[158 \times 111 - (-14)^2]}} \\
&= \frac{(9102 - 140)}{\sqrt{(13986 - 100)(17538 - 196)}} \\
&= \frac{8962}{\sqrt{13886 \times 17342}} \\
&= \frac{8962}{\sqrt{240811012}} \\
&= \frac{8962}{15518.0866} \\
&= 0.5775
\end{aligned}$$

Assumptions of Karl Pearson's coefficient of correlation:

Following are the assumptions of Karl Pearson's coefficient of correlation:

- (i) **Linear relationship:** When the two variables are plotted on a scatter diagram, there will be formed a straight line as result of the points so plotted. The straight line shows a linear relationship between the variables.
- (ii) **Causal Relationship:** There is a cause and effect relationship between the two variables otherwise correlation is meaningless and absurd.
- (iii) **Normality:** The two variables are affected to a great extent and by a large number of independent causes. So that they form a normal distribution.
- (iv) **Proper grouping:** An equal number of pairs helps better.
- (v) **Error of measurement:** It should be reduced to the minimum.

Merits of Karl Pearson's method:

- (i) This method gives a precise and quantitative degree of correlation in one figure of the relationship between two variables. It also shows the direction (i.e., + or -) of relationship.
- (ii) This method helps in obtaining the estimated value of dependent variable from the known value of independent variable along with other information.

Shortcomings/ demerits:

- (i) The coefficient of correlation assumes linear relationship between the two variables whether this assumption is correct or not.
- (ii) The value of coefficient is unduly affected by the extreme items.
- (iii) This method is time consuming to compute the value of coefficient of correlation as compared to other methods.
- (iv) This method may give a misleading picture of the degree of relationship between two variables if the data are not reasonably homogeneous.

Limits: It is relative measure. Its limits are +1 to -1. In other words, it should not be more than +1 or less than -1.

(b) Spearman's Rank Difference Method:

This method was developed by Prof. Charles Spearman to find correlation between

two individuals or ungrouped variables. This method is known as Spearman's Rank Differences Method or Ranking method. It is denoted by ρ (rho). When the data are available in ordinal (rank) form of measurement rather than in interval or ratio form, this type of correlation is useful.

The following formula is used to find Spearman's Rank order correlation:

$$\rho = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$$

Where, D = Difference between the paired ranks

$\sum D^2$ = Sum of squared differences between ranks

N = Number of paired ranks

In a group activity contest of B.Sc. Nursing students, two judges X and Y ranked 10 students as given in following table:

Table 6.31 The Calculation of Rank Order Correlation

Code of Students	Rank given by judge X	Rank given by judge Y	D	D ²
A	3	4	-1	1
B	5	6	-1	1
C	6	5	1	1
D	1	10	-9	81
E	9	8	1	1
F	2	3	-1	1
G	4	2	2	4
H	10	9	1	1
I	7	1	6	36
J	8	7	1	1
				$\sum D^2 = 128$

$$\rho = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$$

Where, $\sum D^2$ = Sum of squared differences between ranks = 128

N = Number of paired ranks = 10

$$\begin{aligned} \rho &= 1 - \frac{6 \times 128}{10(10^2 - 1)} \\ &= 1 - \frac{768}{10(100 - 1)} \\ &= 1 - \frac{768}{10(99)} \\ &= 1 - \frac{768}{990} \\ &= \frac{990 - 768}{990} \\ &= 0.2242 \end{aligned}$$

Let us check our progress-4

(i) Calculate the Pearson's r from following ungrouped data

(a) Table 6.32 Ranks of Two Series

X	55	64	62	69	72	56	65	59	60	64	57	58
Y	52	60	65	56	69	51	56	58	55	58	59	59

(b) Table 6.33 Ranks of Two Series

X	56	64	62	69	72	56	65	59	60	64	59	58
Y	53	60	65	56	69	51	56	58	55	58	60	59

(c) Table 6.34 Ranks of Two Series

X	56	64	62	69	72	56	65	59	60	64	59
Y	53	60	65	56	69	51	56	58	55	58	60

(ii) Calculate the Pearson's r from following grouped data

(a) Table 6.35 Marks obtained in Psychology and Sociology

Marks obtained in Psychology (Y)	Marks obtained in Sociology (X)					
	25-35	35-45	45-55	55-65	65-75	Total
25-35	8	6	-	-	-	14
35-45	5	15	12	-	-	32
45-55	-	10	15	7	-	32
55-65	6	-	5	10	4	25
65-75	-	4	-	-	5	9
Total	19	35	32	17	9	112

(b) Table 6.36 Marks obtained in Microbiology and Nutrition

Marks obtained in Microbiology (Y)	Marks obtained in Nutrition (X)					
	30-40	40-50	50-60	60-70	70-80	Total
30-40	7	3	-	-	-	10
40-50	4	17	13	-	-	34
50-60	-	12	16	7	-	35
60-70	-	-	8	11	2	21
70-80	-	-	-	5	1	6
Total	11	32	37	23	3	106

(c) Table 6.37 Marks obtained in Pathology and Biochemistry

Marks obtained in Pathology (Y)	Marks obtained in Biochemistry (X)					
	25-35	35-45	45-55	55-65	65-75	Total
25-35	8	4	3	-	-	15
35-45	5	18	12	-	-	35
45-55	-	13	15	9	-	37
55-65	-	-	9	12	3	24
65-75	-	-	-	7	2	9
Total	13	35	39	28	5	120

(iii) Calculate co-efficient of correlation by ranking method from following data.

(a) In a class test of two subjects, B.Sc. Nursing students, obtained marks in both subjects as given in following table: 6.35

Table 6.38 Marks obtained in Psychology and Sociology

Code of Students	Marks obtain in Psychology (X)	Marks obtain in Sociology (Y)
A	3	4
B	5	6
C	6	5
D	1	10
E	9	8
F	2	3
G	4	2
H	10	9
I	7	1
J	5	7
K	7	8
L	9	7
M	7	8
N	8	6
O	6	8

(b) Calculate the coefficient of correlation by ranking method:

In a class test of two subjects, B.Sc. Nursing students, obtained marks in both subjects as given in following table:6.36

Table 6.39 Marks obtained in Psychology and Sociology

Code of Students	Marks obtain in Psychology (X)	Marks obtain in Sociology (Y)
A	3	4
B	5	6
C	6	5
D	1	10
E	9	8
F	2	3
G	4	2
H	10	9
I	7	1
J	5	7
K	7	8
L	9	7
M	7	8

(c) In a class test of two subjects, B.Sc. Nursing students, obtained marks in both subjects as given in following table: 640

Table 6.40 Marks obtained in Psychology and Sociology

Code of Students	Marks obtain in Psychology (X)	Marks obtain in Sociology (Y)
A	3	4
B	5	6
C	6	5
D	1	10
E	9	8
F	2	3
G	4	2
H	10	9
I	7	1
J	5	7
K	7	8
L	9	7
M	7	8
N	5	9
O	9	7
P	6	8
Q	7	9

9. Statistical packages and its application

9.1 Statistical packages meaning

A statistical package is the software. It is used to collect, organize, interpret and present numerical information.

1. There are many complexities of calculations for analysis and interpretation in research study. A statistical package is needed to resolve such types of the complexity. It is helpful to bring accuracy in results.
2. ANOVA, correlation covariance, descriptive statistics, exponential smoothing, F-test, moving average, regression analysis, sampling analysis, t-test and z-test are important tools in the MS-excel.
3. Report, descriptive statistics, compare means, general linear model, correlate, regression, classify, data reduction, scale, non-parametric tests and multiple response are some important statistical analysis tools in SPSS.

9.2 Application of statistical packages

Today, as a result of advances in computing, there are many types of software available. These are various statistical packages available according to disciplines. One can easily upload software of his choice or need into his PC. We are describing two packages which are easily available, usable and popular.

9.2.1 Microsoft excel

Microsoft excel is a big worksheet. It can be used for data entry as well as for performing calculations. With the help of "paste function" we can paste any formula. It can be used to create tables and graphs. It also performs statistical calculations. It is a collection of cells. There are 65,000 (rows) X 256 (columns) cells in it. Each row or column can be used to enter data belonging to one category. Data entry in it is like writing on a piece of paper. It supports various data formats. One can choose a data format by formatting the cells.

Once the type of cells is defined it is easy to enter the data without taking care of the format. It can perform usual calculations on the data so entered. It has an insert function (fx) icon. This icon contains many inbuilt functions e.g. sum, count, max/min, standard deviation etc. To perform a calculation, one has to select a function and specify the range of values on which it has to be applied. It has a built-in statistical package for taking us in further details of data analysis. For further analysis, we provide the data and parameters for each analysis. As a result, after the appropriate statistical macro functions, displays the results in an output table. It is used for cross tabulations, pivot tables and the graphical features also. There has a built-in facility to create several types of graphs and charts.

Following tools are available in MS-excel:

- a. ANOVA:** This tool provides different types of variance analysis. It is used to test of the hypothesis.
- b. Correlation/ Covariance:** It calculates the correlation coefficient/ covari-

ance between two variables when measurements on each variable are observed for each of N subjects. It examines each pair of measurement variables to determine whether the two measurement variables tend to move together.

- c. **Descriptive statistics:** It generates a report of univariate statistics for data. It provides information about the central tendency and variability of the data. It describes the data in an interpretable format. It also shows summary statistics like mean, mode, median, standard deviation, skewness, kurtosis and range etc.
- d. **Exponential smoothing:** It predicts a value based on the forecast for the prior period. It is adjusted for the error in that prior forecast. It forecast on the basis of a smoothing constant.
- e. **F-Test:** It performs a two-sample F-test to compare two population variances. It tests those two samples that come from distributions with equal variances.
- f. **Moving average:** It projects values in the forecast period, based on the average value of the variable over a specific number of preceding periods. It forecast on the basis of past figures.
- g. **Regression analysis:** It performs linear regression analysis by using the "least squares" method to fit a line through a set of observations. It analyses how a single dependent variable is affected by the values of one or more independent variables.
- h. **Sampling analysis:** It creates a sample from a population by treating the input range as a population. It infers about a population on the basis of a sample.
- i. **T-test and Z-test:** It determines whether the two samples are likely to have come from distributions with equal population means. It compares two population means when the population variances are known and unknown.

9.2.2 SPSS (Statistical Package for Social Sciences)

It is developed by three graduate students of Stanford University, in the late 1960s. It can take data input from many packages, e.g. dBase (*.dbf), Excel (*.xls), Lotus 123 (*.w*) and others like *.dat and *.txt. It filters the data and performs analysis only in selected cases. After opening a data file, we go to the analyse menu and start working on the statistical aspects of the data.

Following tools are available in Statistical Analysis in SPSS:

- a. **Report:** It performs the procedures as- OLAP (Online Analytical Processing) Cubes, report summaries in rows, summarize procedure and report summaries in columns.
- b. **Descriptive statistics:** It performs the procedures as- frequencies procedure, descriptive procedure, explore procedure, cross-tabs procedure, ratio statistics procedure.
- c. **Compare means:** It performs the procedures as- means procedure, one-sample T -Test procedure, independent samples, t-test procedure, paired-samples

t-test procedure, one-way ANOVA procedure.

- d. **General linear model:** It performs the GLM Univariate procedure.
- e. **Correlate:** It performs the procedures as- bivariate correlations procedure, partial correlations procedure, calculates statistics measuring either similarities or dissimilarities (distances).
- f. **Regression:** It performs the procedures as- linear regression procedure and curve estimation procedure.
- g. **Classify:** It performs the procedures as- K-Means Cluster Analysis procedure, hierarchical cluster analysis procedure, discriminant analysis procedure.
- h. **Data reduction:** It performs factor analysis procedure.
- i. **Scale reliability:** It performs analysis procedure and multidimensional scaling procedure.
- j. **Non-parametric:** It performs procedures as- Chi-Square Test procedure, Binomial Test procedure, Runs Test Tests procedure, One-Sample Kolmogorov-Smirnov Test procedure, Two-Independent-Samples Tests procedure, Tests for Several Independent Samples procedure, Two-Related-Samples Tests procedure, Tests for Several Related Samples procedure
- k. **Multiple response:** It performs procedures as- Define Multiple Response Sets procedure, Multiple Response Frequencies procedure, Multiple Response Cross-tabs procedure

There are a lot of software available in the market. Many of them may be used as shareware, freeware and online pages for statistical calculations. Many of the universities offer statistics online computational resources. The advancement in software technology, results in providing for enhanced statistical functions, enhanced user-friendliness, better graphics and sound technical support.

10. Let us sum up

The statistics is as old as the human society itself. Gottfried Achenwall is known as the "father of statistics". It includes procedures and techniques used to collect process and analyze the numerical data to make inferences. It is used in a singular as well as in plural sense. Statistics has some characteristics and some limitations. Statistics used in every aspect of our daily life including in field of nursing. There are two types of statistics- descriptive statistics and inferential statistics. It is classified into qualitative data and quantitative data. According to its characteristics, the scale/ levels of measurement are classified in nominal level data, ordinal level data, internal level data and ratio level data. There is a need of raw data to into a meaningful system, otherwise it has little meaning. The researcher translates numerical data or facts into more concrete and understandable forms.

The central tendency are important measures and useful in describing the nature of a frequency distribution. The ranges, the variance and standard deviation are most commonly used measures of variability. Normal probability distribution is a probability function that specifies how the values of a variable are distributed and is

based on the law of probability. Tests of significance are a very common and an important term during a statistical process. The researcher has to make sure about the quality of sample, accuracy and good measures which need to find the truth. The researcher has to determine whether the findings of experiments have occurred as a result of a good study or just by chance. Tests of significance are the method of rejecting or supporting the claims based on sample data. The p-value is a function of observed sample results which is being used for testing of statistical hypothesis. A confidence level is an expression of how confident a researcher can be of the data obtained from a sample. There are basically two types of errors: Type I and Type II. When the values of two variables vary in such a way that fluctuations in one are accompanied by fluctuations in the other, these two variables are said to be correlated. Co-efficient of correlation is a measure of the degree of relationship between two sets of measures for the same group of individuals. There are various methods of correlation of variables. The use of correlation is done according to situation and type of data. A statistical package is defined as the software used to collect, organize, interpret and present numerical information. The need of a statistical package arises due to the complexity of calculations involved therein for analysis and inference. It helps to bring accuracy in results.

11. Answers to check our progress

Let us check our progress-1

(i) Define statistics.

Statistics is "Science and arts that gathers and co-ordinates numerus facts within a determined field, treats these mathematically so that the numerical relations between these facts may be displayed clearly and freed from anomalies and chance factors."

- *English and English*

"Statistics comprises the collection, tabulation, presentation and analysis of an aggregate of facts, collected in a methodical manner, without bias and related to a predetermined purpose."

- *W.G. Sutcliffe*

"Statistics is that branch of scientific method which deals with data obtained by counting of measuring properties of population of natural phenomena."

- *M.G. Kendall*

(ii) State the characteristics of statistics.

(A) Statistics consists aggregates of facts

A single and isolated figure or fact cannot be called as statistics. The data (plural sense) are subject matter of statistics not datum. The statistics must consist of aggregate of certain facts. A single and isolated fact or figure like, 160-centimetre height of a nursing student or the 60 marks obtained in applied anatomy of a particular nursing student in mid-term examination are not count as statistics, because these figures are unrelated and incomparable. For a data may amount to statistics it must be in the form of a set or aggregate of certain facts, e.g. height

of 30 students of first year of B.Sc. nursing and marks obtained in applied anatomy of 30 students of first year of B.Sc. nursing in mid-term examination may be count as statistics.

(B) It is affected by multiple causes

It is not easy to study the effects of one factor only by ignoring the effects of other factors in statistics. We have to study the effects of all the factors separately as well as collectively on the phenomenon because effects of the factors can change with change of place, time or situation, e.g. we can say that result of B.Sc. nursing final year examination does not depend on any single factor but collectively on teachers (standard of teachers, teaching methods, teaching aids), students (performance of students in theory, clinical and practicals) and other factors (standard of question papers, methods and standards evaluation).

(C) It should be expressed numerically

There is a main characteristic of statistics that the facts be expressed in numbers, so that statistical processes may carry out. The qualitative expression like he is a young man, he is poor/ rich or illiterate etc. are not considered as statistics but he is 25 years old, his income is Rs. 30000 per month etc. Thus, these data contain numerical figures so that may be called as numerical statement of facts.

(D) It must be enumerated or estimated accurately:

In a particular enquiry, numerical data may be obtained by complete enumeration or by estimation. The use of these both methods depends upon the area and object of enquiry. If the area of enquiry is not very large and complete enumeration is possible, the data collection be made through complete enumeration. On the other side, If the area of enquiry is very large and complete enumeration is not possible or impractical due to certain reasons, the collection of data be made through estimation. It is done by the technique of sampling. e.g. to obtain the knowledge level on nursing care of a particular disease of all nursing students in country. In such case, complete enumeration is not possible or impractical due to certain reasons, the collection of data be made through estimation. But if this enquiry is made on a particular nursing institution, it is possible to obtain data through complete enumeration.

(E) It should be collected in a systemic manner

There should be made a suitable and systematic plan, before collecting the data. The trained investigators should be used to collect data. The data collected in a haphazard manner will lead to difficulties in the process of analysis and show wrong conclusions. If it is not done, in such cases reliability of data gets decreased. So, to get correct results the data must be collected in a precise manner.

(F) It should be collected for a predetermined purpose

We must be clear about the purpose and scope of enquiry before we start the collection of data, If we have no information about its purpose, we may not be able to collect data according to the needs. The definitions of various terms used, unit

and measurement help to achieve the required purpose. Suppose we want to get data on preparedness of vaccination against the covid-19 virus. We have to know about total community health centres, primary health centres, sub centres along with available manpower which may be utilized for the predetermined purpose.

(G) It should have a reasonable standard of accuracy

The estimation of facts and figures cannot be as precise and accurate as actual enumeration or actual measurement. A high degree of accuracy observed in mathematics is not observed in statistics, e.g. while measuring the weight of a patient accuracy will in terms of a fraction of a gram, while measuring the weight of loaded truck accuracy will in terms of a fraction of a kilogram. "What should be the degree of accuracy depends upon the object of enquiry." However, certain standards of accuracy be maintained for drawing meaningful conclusion.

(H) It should be capable of being placed in relation to each other

In any research study, the collection of data is generally done to compare. If the figures collected are not comparable, in that case, they lose a large part of their significance. The comparison is done between homogeneous study and not heterogeneous, e.g. heterogeneous data like average weight of B.Sc. nursing first year students 54 kilograms and distance between college of nursing and clinical area is 10 kilometres can never be placed in relation to each other and compared for analysis and interpretation. It can be concluded that all statistics are numerical data but all numerical data are not statistics unless they satisfy all the essential characteristics of statistics, described above.

(iii) Write the uses of statistics in field of nursing.

Human beings always respond in different manner on issues related to various risk factors, diseases and treatments. So, there is need that all nursing personnel should be able to summarize, interpret and display findings in the right ways.

(a) Enhancing the patient care: Statistics enables nursing personnel to prioritize treatments, determine the needs of a patient and immediate medical attention or follow-up care. The nursing personnel use statistics to identify specific patterns in signs and symptoms. It helps to be ready to any medical changes to patient. It helps to carried out treatment on time by maintaining frequency charts or data sheets to document the timing of medication given to patients.

(b) Supporting evidence-based nursing practices: In current scenario, the nursing profession is going to establish empirical evidence for patient care. For this purpose, there is need that nursing personnel should understand the basic of statistics so that they may be able to read, comprehend and interpret relevant literature.

(c) Comparing the options available for nursing practice: Nursing personnel use statistics to determine the effectiveness of the options of the nursing interventions of their choice. In some settings, nursing personnel use their knowledge related to statistics, when have to make several decisions during critical care.

(d) Healthcare has become more complex and challenging: Nursing has become more challenging due to electronic information and published literature that exist. There is always a need of new knowledge to improve patient care. The nursing personnel have to continue learning during the practice.

(e) There is demand of knowledgeable and competent leadership: Formal nursing education is an important component of the commitment to become a leader in the nursing profession. For one to become a nursing leader, they must be able to keep up with the current trends, be able to solve nursing related problems and know about new technology that can improve patient care.

(iv) Enumerate the type of statistics.

There are two types of statistics-

- a. Descriptive statistics
- b. Inferential statistics

Let us check our progress-2

(i) Table 6.41 Differentiate between Continuous and Discrete Series of Data

S.No.	Base of Comparison	Continuous Series	Discrete Series
1.	Meaning	It is one that falls on a continuous sequence/series. It has not clear spaces between values.	It is one that has clear spaces between values.
2.	Nature	It is measurable	It is countable
3.	Values	It can take any value in some value within range.	It can take only distinct or separate values.
4.	Graphical Representation	It shows histogram in graphical representation.	It shows bar graph in graphical representation.
5.	Tabulation	Its table shows grouped frequency distribution.	Its table shows ungrouped frequency distribution.
6.	Classification	It contains over lapping or mutually exclusive classifications, e.g. 5-10, 10-15, 15-20...etc.	It contains non-overlapping or mutually inclusive classifications e.g. 5-9, 10-14, 15-19...etc.
7.	Function graph	It shows connected points	It shows isolated points
8.	Example	Market price of a product	Days of the week

It is clear that continuous data has any value from a given range (without breaks) whereas the discrete data has a certain number of isolated values.

(ii) Describe level of measurement.

According to its characteristics, the scale/levels of measurement are classified in following levels-

(a) Nominal level data- It is first level of data. It is assigned only name and frequency of each category is counted. There is no statistically relationship. Only mode is calculated as measurement of central tendency. When two categories of data occur in the same frequency, it is called bi-modal. When more than two categories of data

occur in the same frequency, it is called multi-modal. It is used in those situations where counting is the only feasible method of quantification.

(b) Ordinal level data- It is the second level of data. It can assign category and rank order, e.g. high, medium, low or from highest to lowest etc. Ordinal level data has no absolute value. When ranking is done, it says equally rank spaces but value wise they are not equally spaced.

(c) Interval level data- It is the third level of data. It has no absolute zero and therefore no fixed beginning point. It has category, rank and equal spaces or intervals or relative distance between the units and a rule of order. It measures mean as central tendency.

(d) Ratio level data- It is the fourth level of data. It has category, rank and equal spaces or intervals or relative distance between the units and a rule of order. Besides it, it has absolute zero and therefore fixed beginning point. It is most precise and highest level of data. It has the numerical values are real number and can be added, subtracted, multiplied, divided and expressed in ratio relationship.

Let us check our progress-3

1. Calculate the mean for the following frequency distribution.

(a) 75 56 66 35 56 81 67 54 71 64 37 68 66 32 70 48 71 57 70 68 63 38 58 78 71
Formula for the mean (M), when data are ungrouped

$M = \Sigma X / N$ (Here, M stands for mean, the symbol Σ stands for 'sum of', X stands for score or other measure and N is the number of measures in the series)

Now, we calculate the mean for the ungrouped scores given in question.

$M = \Sigma X / N$ (Here, $\Sigma X = 1520$ and $N = 25$)

$M = 1520 / 25 = 60.80$

(b) 62 55 59 63 74 65 46 48 77 46 62 52 83 63 73 87 71 41 51 43 62 52 69 78 69
Formula for the mean (M), when data are ungrouped

$M = \Sigma X / N$ (Here, M stands for mean, the symbol Σ stands for 'sum of', X stands for score or other measure and N is the number of measures in the series)

Now, we calculate the mean for the ungrouped scores given in question.

$M = \Sigma X / N$ (Here, $\Sigma X = 1551$ and $N = 25$)

$M = 1551 / 25 = 62.04$

(c) 17 15 16 13 15 18 16 15 17 16 13 16 16 13 17 14 17 15 17 16 16 13 15 17 17
Formula for the mean (M), when data are ungrouped

$M = \Sigma X / N$ (Here, M stands for mean, the symbol Σ stands for 'sum of', X stands for score or other measure and N is the number of measures in the series)

Now, we calculate the mean for the ungrouped scores given in question.

$M = \Sigma X / N$ (Here, $\Sigma X = 390$ and $N = 25$)

$M = 390 / 25 = 15.6$

2. Calculate the mean for the following frequency distribution.**(a) Table 6.42 The Grouped Data into Frequency Distribution**

S.No.	Class Intervals	Mid-Point (X)	Frequencies (f)	(f X)
(1)	(2)	(3)	(4)	(5)
1.	180-185	182	1	182
2.	175-180	177	2	354
3.	170-175	172	4	688
4.	165-170	167	5	835
5.	160-165	162	8	1296
6.	155-160	157	10	1570
7.	150-155	152	6	912
8.	145-150	147	4	588
9.	140-145	142	4	568
10.	135-140	137	2	274
11.	130-135	132	3	396
12.	125-130	127	1	127
			N=50	7790

Formula for the mean (M), when data are grouped into a frequency distribution
 $M = \frac{\sum fX}{N}$ (Here, M stands for mean, the symbol Σ stands for 'sum of', fX stands for the sum of the midpoints multiplied by their frequencies and N is the number of measures in the series)

Now, we calculate the mean for the grouped scores given in question.

First, we calculate midpoint (X) for each class interval. The scores grouped into class intervals lose their own identity, so they are represented by the particular interval on which they fall. The fX is calculated by multiplying the midpoint of each class interval by their respective frequencies. The fX is added and divided by N. We will obtain mean as a result of this process.

$M = \frac{\sum fX}{N}$ (Here, $\sum fX = 7790$ and $N = 50$)

$$= 7790/50$$

$$= 155.8$$

(b) Table 6.43 The Grouped Data into Frequency Distribution

S.No.	Class Intervals	Mid-Point (X)	Frequencies (f)	(f X)
(1)	(2)	(3)	(4)	(5)
1.	175-180	177	1	177
2.	170-175	172	2	344

Contd...

S.No.	Class Intervals	Mid-Point (X)	Frequencies (f)	(f X)
3.	165-170	167	4	668
4.	160-165	162	5	810
5.	155-160	157	8	1256
6.	150-155	152	10	1520
7.	145-150	147	6	882
8.	140-145	142	4	568
9.	135-140	137	4	548
10.	130-135	132	2	264
11.	125-130	127	3	381
12.	120-125	122	1	122
			N=50	7540

Formula for the mean (M), when data are grouped into a frequency distribution $M = \Sigma fX / N$ (Here, M stands for mean, the symbol Σ stands for 'sum of', fX stands for the sum of the midpoints multiplied by their frequencies and N is the number of measures in the series)

Now, we calculate the mean for the grouped scores given in question.

First, we calculate midpoint (X) for each class interval. The scores grouped into class intervals lose their own identity, so they are represented by the particular interval on which they fall. The fX is calculated by multiplying the midpoint of each class interval by their respective frequencies. The fX is added and divided by N. We will obtain mean as a result of this process.

$M = \Sigma fX / N$ (Here, $\Sigma fX = 7540$ and $N = 50$)

$= 7540 / 50$

$= 150.8$

(c) Table 6.44 The Grouped Data into Frequency Distribution

S.No.	Class Intervals	Mid-Point (X)	Frequencies (f)	(f X)
1.	170-175	172	1	172
2.	165-170	167	2	334
3.	160-165	162	4	648
4.	155-160	157	5	785
5.	150-155	152	8	1216
6.	145-150	147	10	1470
7.	140-145	142	6	852
8.	135-140	137	4	548

Contd...

S.No.	Class Intervals	Mid-Point (X)	Frequencies (f)	(f X)
9.	130-135	132	4	528
10.	125-130	127	2	254
11.	120-125	122	3	366
12.	115-120	117	1	117
			N=50	7290

Formula for the mean (M), when data are grouped into a frequency distribution
 $M = \Sigma fX / N$ (Here, M stands for mean, the symbol Σ stands for 'sum of', fX stands for the sum of the midpoints multiplied by their frequencies and N is the number of measures in the series)

Now, we calculate the mean for the grouped scores given in question.

First, we calculate midpoint (X) for each class interval. The scores grouped into class intervals lose their own identity, so they are represented by the particular interval on which they fall. The fX is calculated by multiplying the midpoint of each class interval by their respective frequencies. The fX is added and divided by N. We will obtain mean as a result of this process.

$$M = \Sigma fX / N \text{ (Here, } \Sigma fX = 7290 \text{ and } N = 50)$$

$$= 7290 / 50$$

$$= 145.8$$

3. Calculate the median for the following frequency distribution.

(a) 75 56 66 35 56 81 67 54 71 64 37 68 66 32 70 48 71 57 70 68 63 38 58 78 71

To calculate the median from the above test scores, we have to arrange these twenty five scores in order of size:

32 35 37 38 48 54 56 56 57 58 63 64 66 66 67 68 68 70 70 71 71 71 75 78 81

Here, midpoint is the 66, twelve scores are above 66 and twelve scores are below 66. Hence, median is 66.

(b) 62 55 59 63 74 65 46 48 77 46 62 52 83 63 73 87 71 41 51 43 62 52 69 78 69

To calculate the median from the above test scores, we have to arrange these twenty five scores in order of size:

41 43 46 46 48 51 52 52 55 59 62 62 62 63 63 65 69 69 71 73 74 77 78 83 87

Here, midpoint is the 62, twelve scores are above 62 and twelve scores are below 62. Hence, median is 62.

(c) 17 15 16 13 15 18 16 15 17 16 13 16 16 13 17 14 17 15 17 16 16 13 15 17 17

To calculate the median from the above test scores, we have to arrange these twenty five scores in order of size:

13 13 13 13 14 15 15 15 15 15 16 16 16 16 16 16 16 16 17 17 17 17 17 17 18

Here, midpoint is the 16, twelve scores are above 16 and twelve scores are below 16. Hence, median is 16.

4. Calculate the median for the following frequency distribution.

(a) Imagine that in the statistics subject, students got the marks in a class test as given below:

18 19 12 11 14 17 18 10

To calculate the median from the above test scores, we have to arrange these eight scores in order of size: 10 11 12 14 17 18 18 19

We use formula to find out median from ungrouped scores is given below:

Median (Md.) = the $(N+1)/2^{\text{th}}$ measure in order of size

$$= 8+1/2$$

$$= 4.5^{\text{th}} \text{ score in size of order}$$

Here, the 4.5th score in size of order is 4th score (14) & 5th score (17). The midpoint is the 15.5 because it is higher limit of score 14 and lower limit of score 17.

Hence, median is 15.5.

(b) Imagine that in the Statistics subject, students got the marks in a class test as given below:

18 19 12 11 14 16 18 10

To calculate the median from the above test scores, we have to arrange these eight scores in order of size: 10 11 12 15 16 18 18 19

We use formula to find out median from ungrouped scores is given below:

Median (Md.) = the $(N+1)/2^{\text{th}}$ measure in order of size

$$= 8+1/2$$

$$= 4.5^{\text{th}} \text{ score in size of order}$$

Here, the 4.5th score in size of order is 4th score (15) & 5th score (16). The midpoint is the 15.5.

Hence, median is 15.5.

(c) Imagine that in the Statistics subject, students got the marks in a class test as given below:

28 29 12 11 14 27 28 10 30 31

To calculate the median from the above test scores, we have to arrange these ten scores in order of size: 10 11 12 14 27 28 28 29 30 31

We use formula to find out median from ungrouped scores is given below:

Median (Md.) = the $(N+1)/2^{\text{th}}$ measure in order of size

$$= 10+1/2$$

$$= 5.5^{\text{th}} \text{ score in size of order}$$

Here, the 5.5th score in size of order is 5th score (27) & 6th score (28). The midpoint is the 27.5. Hence, median is 27.5.

5. Calculate the median for the following frequency distribution.

(a) When scores are available in a continuous series and grouped into a frequency distribution. The median will be 50% point in distribution. We may find out median with the help of following process.

Table 6.45 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Frequency (f)	Counting 50% point
(1)	(2)	(3)	(4)
1.	180-185	1	
2.	175-180	2	
3.	170-175	4	
4.	165-170	5	
5.	160-165	8	
6.	155-160	10	
7.	150-155	6	20
8.	145-150	4	14
9.	140-145	4	10
10.	135-140	2	6
11.	130-135	3	4
12.	125-130	1	1
N=50			

In the above Table 6.45, to find out the median, it is taken 50% of our scores ($N/2=25$).

We begin counting from lower side or small score.

We start at the class interval 125-130 and adding up scores serially, we reach the class interval 150-155. Here, we find 20 frequencies, it is short to 25 frequencies, which are necessary to find out the median ($N/2=25$). The next class interval is 155-160 which contain 10 frequencies. It is assumed that the median will present in this class interval. To fulfill the requirement of five frequencies, we take $5/10 \times 5$ (5=needed frequencies, 10=frequencies of next class interval and 5=the length of the interval). By this calculation, we get 2.5 and add it in the 154.5, the beginning of the class interval 155-160. We get the median at $154.5 + 2.5=157$.

Formula for calculating the median (Mdn), when data are grouped into frequency distribution

$$\text{Mdn} = l + \{(N/2 - F)/f_m\}i$$

(Here, Mdn stands for median, the l = exact lower limit of the class interval upon which the median lies, $N/2$ = one half of the total number of scores, F = sum of the scores on all intervals below l , f_m =frequency (number of scores) within the interval upon which the median fall, i =length of the class interval)

We will calculate the median from the data given in the Table 6.45

$$\text{Mdn} = l + \{(N/2 - F)/f_m\}i$$

(Here, the l = 154.5, $N/2$ = $50/2=25$, F = 20, f_m = 10, i = 5)

$$\text{Mdn} = 154.5 + \{(50/2 - 20)/10\} \times 5$$

$$\begin{aligned}
&= 154.5 + \{(25-20)/10\} \times 5 \\
&= 154.5 + (5/10) \times 5 \\
&= 154.5 + (0.5) \times 5 \\
&= 154.5 + 2.5
\end{aligned}$$

$$\text{Mdn} = 157$$

(b) When scores are available in a continuous series and grouped into a frequency distribution. The median will be 50% point in distribution. We may find out median with the help of following process.

Table 6.46 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Frequencies (f)	Counting 50% point
(1)	(2)	(4)	(5)
1.	175-180	1	
2.	170-175	2	
3.	165-170	4	
4.	160-165	5	
5.	155-160	8	
6.	150-155	10	
7.	145-150	6	20
8.	140-145	4	14
9.	135-140	4	10
10.	130-135	2	6
11.	125-130	3	4
12.	120-125	1	1
N=50			

In the above table 6.43, to find out the median, it is taken 50% of our scores ($N/2=25$). We begin counting from lower side or small score.

We start at the class interval 120-125 and adding up scores serially, we reach the class interval 145-150. Here, we find 20 frequencies, it is short to 25 frequencies, which are necessary to find out the median ($N/2=25$). The next class interval is 150-155 which contain 10 frequencies. It is assumed that the median will present in this class interval. To fulfil the requirement of five frequencies, we take $5/10 \times 5$ (5=needed frequencies, 10=frequencies of next class interval and 5=the length of the interval). By this calculation, we get 2.5 and add it in the 149.5, the beginning of the class interval 150-155. We get the median at $149.5 + 2.5=152$.

Formula for calculating the median (Mdn), when data are grouped into frequency distribution

$$\text{Mdn} = l + \{(N/2 - F)/f_m\}i$$

(Here, Mdn stands for median, the l = exact lower limit of the class interval upon

which the median lies, $N/2$ = one half of the total number of scores, F = sum of the scores on all intervals below l , f_m = frequency (number of scores) within the interval upon which the median fall, i = length of the class interval)

We will calculate the median from the data given in the Table 6.43

$$\text{Mdn} = l + \{(N/2 - F)/f_m\}i$$

(Here, the $l = 149.5$, $N/2 = 50/2 = 25$, $F = 20$, $f_m = 10$, $i = 5$)

$$\begin{aligned}\text{Mdn} &= 149.5 + \{(50/2 - 20)/10\} \times 5 \\ &= 149.5 + \{(25 - 20)/10\} \times 5 \\ &= 149.5 + (5/10) \times 5 \\ &= 149.5 + (0.5) \times 5 \\ &= 149.5 + 2.5\end{aligned}$$

$$\text{Mdn} = 152$$

(c) When scores are available in a continuous series and grouped into a frequency distribution. The median will be 50% point in distribution. We may find out median with the help of following process.

Table 6.47 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Frequencies (f)	Counting 50% point
(1)	(2)	(4)	(5)
1.	170-175	1	
2.	165-170	2	
3.	160-165	4	
4.	155-160	5	
5.	150-155	8	
6.	145-150	10	
7.	140-145	6	20
8.	135-140	4	14
9.	130-135	4	10
10.	125-130	2	6
11.	120-125	3	4
12.	115-120	1	1
N=50			

In the above Table 6.47, to find out the median, it is taken 50% of our scores ($N/2 = 25$).

We begin counting from lower side or small score.

We start at the class interval 115-120 and adding up scores serially, we reach the class interval 140-145. Here, we find 20 frequencies, it is short to 25 frequencies, which are necessary to find out the median ($N/2 = 25$). The next class interval is 145-150 which contain 10 frequencies. It is assumed that the median will present in this class interval. To fulfil the requirement of five frequencies, we take $5/10 \times 5$

(5=needed frequencies, 10=frequencies of next class interval and 5=the length of the interval). By this calculation, we get 2.5 and add it in the 144.5, the beginning of the class interval 150-155. We get the median at $144.5 + 2.5 = 147$.

Formula for calculating the median (Mdn), when data are grouped into frequency distribution

$$\text{Mdn} = l + \{(N/2 - F)/f_m\}i$$

(Here, Mdn stands for median, the l = exact lower limit of the class interval upon which the median lies, $N/2$ = one half of the total number of scores, F = sum of the scores on all intervals below l , f_m =frequency (number of scores) within the interval upon which the median fall, i =length of the class interval)

We will calculate the median from the data given in the Table 6.47

$$\text{Mdn} = l + \{(N/2 - F)/f_m\}i$$

(Here, the l = 144.5, $N/2$ = $50/2=25$, F = 20, f_m = 10, i = 5)

$$\begin{aligned}\text{Mdn} &= 144.5 + \{(50/2 - 20)/10\} \times 5 \\ &= 144.5 + \{(25 - 20)/10\} \times 5 \\ &= 144.5 + (5/10) \times 5 \\ &= 144.5 + (0.5) \times 5 \\ &= 144.5 + 2.5\end{aligned}$$

$$\text{Mdn} = 147$$

6. Calculate the mode for the following frequency distribution.

(a) 75 56 66 35 56 81 67 54 71 64 37 68 66 32 70 48 71 57 70 68 63 38 58 78 71

Mode is the most frequently occurring value in that set. To determine the mode, the scores are arranged in an order and then counted each one. The most frequently occurring value is the mode.

32 35 37 38 48 54 56 56 57 58 63 64 66 66 67 68 68 70 70 71 71 71 75 78 81 is a set of scores.

In this set of scores, only the score 71, occurs three times so it is the mode of this set of scores.

(b) 62 55 59 63 74 65 46 48 77 46 62 52 83 63 73 87 71 41 51 43 62 52 69 78 69

Mode is the most frequently occurring value in that set. To determine the mode, the scores are arranged in an order and then counted each one. The most frequently occurring value is the mode.

41 43 46 46 48 51 52 52 55 59 62 62 62 63 63 65 69 69 71 73 74 77 78 83 87 is a set of scores.

In this set of scores, only the score 62, occurs three times, so it is the mode of this set of scores.

(c) 17 15 16 13 15 18 16 15 17 16 13 16 16 13 17 14 17 15 17 16 16 13 15 17 17

Mode is the most frequently occurring value in that set. To determine the mode, the scores are arranged in an order and then counted each one. The most frequently occurring value is the mode.

13 13 13 13 14 15 15 15 15 15 16 16 16 16 16 16 16 16 17 17 17 17 17 17 17 18 is a set of scores.

In this set of scores, the score 16 and 17 both, occurs seven times so there is present of two modes. This condition is called, bimodal distribution.

7. Calculate the mode for the following frequency distribution.

(a) When data are grouped into the frequency distribution. We can take the crude or empirical mode from mid-point of that interval, which contains the largest frequency. We can refer Table 6.45, in this table; the class interval 155-160 contains the largest frequency (10). Its mid-point is (157 or 156.5), hence 156.5 is the crude mode of this grouped frequency distribution.

Table 6.48 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Mid Point (X)	(f)
(1)	(2)	(3)	(4)
1.	180-185	182	1
2.	175-180	177	2
3.	170-175	172	4
4.	165-170	167	5
5.	160-165	162	8
6.	155-160	157	10
7.	150-155	152	6
8.	145-150	147	4
9.	140-145	142	4
10.	135-140	137	2
11.	130-135	132	3
12.	125-130	127	1
N=50			

After calculating the crude mode, now we will calculate the true mode. The true mode is the point of greatest concentration in the frequency distribution. It is calculated by the following formula:

$$Mode = l + \left(\frac{fm - f_1}{2fm - f_1 - f_2} \right) \times i$$

Here, l = Lower limit of the modal class interval (having maximum frequency)
 fm = Frequency of modal class interval
 f_1 = Frequency of class interval preceding the modal class interval
 f_2 = Frequency of class interval following the modal class interval
 i = Width of the modal class interval

Now, we will practice of the formula with the table 6.48. In this table, the maximum frequency is 10, which lies in class interval 154.5-159.5.

Here, l = Lower limit of the modal class interval (having maximum frequency) = 154.5
 fm = Frequency of modal class interval = 10
 f_1 = Frequency of class interval preceding the modal class interval = 8
 f_2 = Frequency of class interval following the modal class interval = 6
 i = Width of the modal class interval = 5

Let us put in formula

$$\begin{aligned}
 \text{Mode} &= l + \left(\frac{fm - f_1}{2fm - f_1 - f_2} \right) \times i \\
 &= 154.5 + \{(10-8)/(2 \times 10 - 8 - 6)\} \times 5 \\
 &= 154.5 + \{2/(20-8-6)\} \times 5 \\
 &= 154.5 + 2/6 \times 5 \\
 &= 154.5 + 10/6 \\
 &= 154.5 + 1.67 \\
 \text{Mode} &= 156.17
 \end{aligned}$$

When the frequency distribution is symmetrical, the formula is used to determine the mode.
 Mode=3 Median-2 Mean

(b) When data are grouped into the frequency distribution. We can take the crude or empirical mode from mid-point of that interval, which contains the largest frequency. We can refer Table 6.46, in this table, the class interval 150-155 contains the largest frequency (10). Its mid-point is (152 or 151.5), hence 151.5 is the crude mode of this grouped frequency distribution.

Table 6.49 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Mid Point (X)	(f)
(1)	(2)	(3)	(4)
1.	175-180	177	1
2.	170-175	172	2
3.	165-170	167	4
4.	160-165	162	5
5.	155-160	157	8
6.	150-155	152	10
7.	145-150	147	6
8.	140-145	142	4
9.	135-140	137	4
10.	130-135	132	2
11.	125-130	127	3
12.	120-125	122	1
N=50			

After calculating the crude mode, now we will calculate the true mode. The true mode is the point of greatest concentration in the frequency distribution. It is calculated by the following formula:

$$\text{Mode} = l + \left(\frac{fm - f_1}{2fm - f_1 - f_2} \right) \times i$$

Here, l= Lower limit of the modal class interval (having maximum frequency)
 fm= Frequency of modal class interval
 f₁ = Frequency of class interval preceding the modal class interval
 f₂ = Frequency of class interval following the modal class interval
 i = Width of the modal class interval

Now, we will practice of the formula with the table 6.46 In this table, the maximum frequency is 10, which lies in class interval 149.5-154.5.

Here, l = Lower limit of the modal class interval (having maximum frequency) = 149.5

f_m = Frequency of modal class interval = 10

f_1 = Frequency of class interval preceding the modal class interval = 8

f_2 = Frequency of class interval following the modal class interval = 6

i = Width of the modal class interval = 5

Let us put in formula

$$\begin{aligned}
 \text{Mode} &= l + \left(\frac{f_m - f_1}{2f_m - f_1 - f_2} \right) \times i \\
 &= 149.5 + \{(10-8)/(2 \times 10 - 8 - 6)\} \times 5 \\
 &= 149.5 + \{2/(20-8-6)\} \times 5 \\
 &= 149.5 + 2/6 \times 5 \\
 &= 149.5 + 10/6 \\
 &= 149.5 + 1.67 \\
 \text{Mode} &= 151.17
 \end{aligned}$$

When the frequency distribution is symmetrical, the formula is used to determine the mode.

Mode = 3 Median - 2 Mean

(c) When data are grouped into the frequency distribution. We can take the crude or empirical mode from mid-point of that interval, which contains the largest frequency. We can refer Table 6.47, in this table, the class interval 145-150 contains the largest frequency (10). Its mid-point is (147 or 144.5), hence 144.5 is the crude mode of this grouped frequency distribution.

Table 6.50 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Mid Point (X)	(f)
(1)	(2)	(3)	(4)
1.	170-175	172	1
2.	165-170	167	2
3.	160-165	162	4
4.	155-160	157	5
5.	150-155	152	8
6.	145-150	147	10
7.	140-145	142	6
8.	135-140	137	4
9.	130-135	132	4
10.	125-130	127	2
11.	120-125	122	3
12.	115-120	117	1
			N=50

After calculating the crude mode, now we will calculate the true mode. The true mode is the point of greatest concentration in the frequency distribution. It is calculated by the following formula:

Here, l = Lower limit of the modal class interval (having maximum frequency)

fm = Frequency of modal class interval

$$Mode = l + \left(\frac{fm - f_1}{2fm - f_1 - f_2} \right) \times i$$

f_1 = Frequency of class interval preceding the modal class interval

f_2 = Frequency of class interval following the modal class interval

i = Width of the modal class interval

Now, we will practice of the formula with the table 6.47. In this table, the maximum frequency is 10, which lies in class interval 144.5-149.5.

Here, l = Lower limit of the modal class interval (having maximum frequency) = 144.5

fm = Frequency of modal class interval = 10

f_1 = Frequency of class interval preceding the modal class interval = 8

f_2 = Frequency of class interval following the modal class interval = 6

i = Width of the modal class interval = 5

Let us put in formula

$$\begin{aligned}
 Mode &= l + \left(\frac{fm - f_1}{2fm - f_1 - f_2} \right) \times i \\
 &= 144.5 + \{ (10-8) / (2 \times 10 - 8 - 6) \} \times 5 \\
 &= 144.5 + \{ 2 / (20 - 8 - 6) \} \times 5 \\
 &= 144.5 + 2 / 6 \times 5 \\
 &= 144.5 + 10 / 6 \\
 &= 144.5 + 1.67 \\
 Mode &= 146.17
 \end{aligned}$$

8. Calculate the standard deviation from the following data. (Take the help of deviation scores)

(a) 26, 28, 29, 25, 24, 27, 23, 26

The formula used for calculating standard deviation for ungrouped set of data is:

a. Calculate the mean of set of scores is $= M = \Sigma X / N = 208 / 8 = 26$

b. Find out the value of x , the mean (26) is subtracted from raw scores. The values that are below the mean have negative deviations/ discrepancies and values above the mean are positive.

c. Square each deviation

d. Sum these squares and get sum of squares value. ($\Sigma x^2 = 28$)

Table 6.51 Computing SD from the ungrouped set of Scores

S.No.	Scores(X)	Scores-Mean (X-M=x) (Deviation from the Mean) (x)	Squared of Deviation(x ²)
1.	26	26-26= 0	0
2.	28	28-26= 2	4
3.	29	29-26= 3	9
4.	25	25-26= -1	1
5.	24	24-26= -2	4
6.	27	27-26= 1	1
7.	23	23-26= -3	9
8.	26	26-26= 0	0
			$\Sigma x^2 = 28$

f. Apply the formula for calculation of standard deviation (To get the standard deviation, we take the square root of the variance because we have squared the deviations earlier).

$$\begin{aligned}
 SD &= \sqrt{\Sigma x^2 / n-1} \text{ (Here, } \Sigma x^2 = 28 \text{ and } n=8) \\
 &= \sqrt{28 / 8-1} \\
 &= \sqrt{4}
 \end{aligned}$$

$$SD = 2$$

(b) 20 25 28 25 24 26 24 26 29 23

a. Calculate the mean of set of scores is $=M=\Sigma X/N=250/10=25$

b. Find the value of x, the mean (25) is subtracted from raw scores. The values that are below the mean have negative deviations/ discrepancies and values above the mean are positive.

c. Square each deviation

d. Sum these squares and get sum of squares value. ($\Sigma x^2 = 58$)

Table 6.52 Computing SD from the ungrouped set of Scores

S.No.	Scores(X)	Scores-Mean (X-M=x) (Deviation from the Mean) (x)	Squared of Deviation(x ²)
1.	20	20-25= -5	25
2.	25	25-25= 0	0
3.	28	28-25= 3	9
4.	25	25-25= 0	0
5.	24	24-25= -1	1
6.	26	26-25= 1	1

S.No.	Scores(X)	Scores-Mean (X-M=x) (Deviation from the Mean) (x)	Squared of Deviation(x ²)
7.	24	24-25= -1	1
8.	26	26-25= 1	1
9.	29	29-25=4	16
10.	23	23-25=-2	4
			$\Sigma x^2 = 58$

f. Apply the formula for calculation of standard deviation (To get the standard deviation, we take the square root of the variance because we have squared the deviations earlier).

$$SD = \sqrt{\Sigma x^2 / n-1} \text{ (Here, } \Sigma x^2 = 58 \text{ and } n=10)$$

$$= \sqrt{58/10-1}$$

$$= \sqrt{6.4444}$$

$$SD = 2.5385$$

(C) 16 18 19 15 14 17 13 16

a. Calculate the mean of set of scores is $=M=\Sigma X/N=128/8=16$

b. Find the value of x, the mean (16) is subtracted from raw scores. The values that are below the mean have negative deviations/ discrepancies and values above the mean are positive.

c. Square each deviation

d. Sum these squares and get sum of squares value. ($\Sigma x^2 = 28$)

Table 6.53 Computing SD from the Ungrouped set of Scores

S.No.	Scores(X)	Scores-Mean (X-M=x) (Deviation from the Mean) (x)	Squared of Deviation(x ²)
1.	16	16-16= 0	0
2.	18	18-16= 2	4
3.	19	19-16= 3	9
4.	15	15-16= -1	1
5.	14	14-16= -2	4
6.	17	17-16= 1	1
7.	13	13-16= -3	9
8.	16	16-16= 0	0
			$\Sigma x^2 = 28$

g. Apply the formula for calculation of standard deviation (To get the standard deviation, we take the square root of the variance because we have squared the deviations earlier).

$$SD = \sqrt{\Sigma x^2 / n-1} \text{ (Here, } \Sigma x^2 = 28 \text{ and } n=8)$$

$$= \sqrt{28/8-1}$$

$$= \sqrt{4}$$

$$SD = 2$$

9. Calculate the standard deviation from the following ungrouped data. (Take the help of raw scores)

(a) 26, 28, 29, 25, 21, 27, 22, 26

Table 6.54 Computing SD from the ungrouped set of Scores

S.No.	Scores(X)	Squared of Scores(X ²)
1.	26	676
2.	28	784
3.	29	841
4.	25	625
5.	21	441
6.	27	729
7.	22	484
8.	26	676
ΣX=204		ΣX² =5256

a. Calculate the sum of scores is = ΣX=204

b. Calculate the sum of squared scores is = ΣX² = 5256

d. Apply the formula for calculation of standard deviation

$$\text{Standard Deviation} = (\sigma) = \sqrt{N \cdot \Sigma X^2 - (\Sigma X)^2 / (N-1)}$$

$$= \sqrt{8 \times 5256 - (204)^2 / (8-1)}$$

$$= \sqrt{(42048 - 41616) / (8-1)}$$

$$= \sqrt{432/7}$$

$$= 20.7846/7$$

$$= 2.9692$$

(b) 16, 18, 19, 15, 11, 17, 12, 16, 18

Table 6.55 Computing SD from the ungrouped set of Scores

S.No.	Scores(X)	Squared of Scores(X ²)
1.	16	256
2.	18	324
3.	19	361
4.	15	225
5.	11	221
6.	17	289
7.	12	144
8.	16	256
9.	18	324
ΣX=142		ΣX² =2400

- a. Calculate the sum of scores is = $\Sigma X = 142$
- b. Calculate the sum of squared scores is = $\Sigma X^2 = 2400$
- c. Apply the formula for calculation of standard deviation

$$\begin{aligned}
 \text{Standard Deviation} &= (\sigma) = \sqrt{N \cdot \Sigma X^2 - (\Sigma X)^2 / (N-1)} \\
 &= \sqrt{9 \times 2400 - (142)^2 / (9-1)} \\
 &= \sqrt{21600 - 20164 / 8} \\
 &= \sqrt{1436 / 8} \\
 &= 37.8945 / 8 \\
 &= 4.7368
 \end{aligned}$$

(c) 16, 18, 19, 15, 11, 17, 12, 16, 18, 17

Table 6.56 Computing SD from the Ungrouped set of Scores

S.No.	Scores(X)	Squared of Scores(X ²)
1.	16	256
2.	18	324
3.	19	361
4.	15	225
5.	11	221
6.	17	289
7.	12	144
8.	16	256
9.	18	324
10.	17	289
$\Sigma X = 159$		$\Sigma X^2 = 2689$

- a. Calculate the sum of scores is = $\Sigma X = 159$
- b. Calculate the sum of squared scores is = $\Sigma X^2 = 2689$
- d. Apply the formula for calculation of standard deviation

$$\begin{aligned}
 \text{Standard Deviation} &= (\sigma) = \sqrt{N \cdot \Sigma X^2 - (\Sigma X)^2 / (N-1)} \\
 &= \sqrt{10 \times 2689 - (159)^2 / (10-1)} \\
 &= \sqrt{26890 - 25281 / 9} \\
 &= \sqrt{1609 / 9} \\
 &= 40.1123 / 9 \\
 &= 4.4569
 \end{aligned}$$

10. Calculate the standard deviation from the following grouped set of scores. (Take the help of assumed mean)

(a)

Table 6.57 The data grouped into a frequency distribution

S.No.	Class Intervals	Mid-Point(x)	(f)	(x')	(fx')	(fx' ²)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.	175-180	177	1	5	5	25
2.	170-175	172	2	4	8	32
3.	165-170	167	4	3	12	36
4.	160-165	162	5	2	10	20
5.	155-160	157	8	1	8 (+43)	8
6.	150-155	152	10	0	0	0
7.	145-150	147	5	-1	-5	5
8.	140-145	142	4	-2	-8	16
9.	135-140	137	4	-3	-12	36
10.	130-135	132	2	-4	-8	32
11.	125-130	127	3	-5	-15	75
12.	120-125	122	2	-6	-12 (-60)	72
			N=50	Σfx' = -17		
				Σfx'² = 357		

The formula used for calculating Standard Deviation for grouped set of data with the help of assumed mean deviation scores is:

$$\text{Standard Deviation} = (\sigma) = i / N [\sqrt{N \cdot \Sigma fx'^2 - (\Sigma fx')^2}]$$

(i= Width of the class-interval, x\= Deviation of the raw measures/scores from the assumed mean, N= The number of scores in the distribution)

a. We take the set of grouped set of data in Table 6.57:

b. Calculate deviation of raw data from the assumed mean= x'

c. Calculate the sum of deviation of raw data from the assumed mean is =fx' = -17

c. Calculate the sum of squared deviation of raw data from the assumed mean =
fx'² = 357

d. Apply the formula for calculation of

$$\text{Standard Deviation} = (\sigma) = i / N [\sqrt{N \cdot \Sigma fx'^2 - (\Sigma fx')^2}]$$

$$\Sigma fx' = -17 \quad \Sigma fx'^2 = 357$$

(i= Width of the class-interval=5, x' = Deviation of the raw measures/scores from the assumed mean, sum of squared deviation of raw data from the assumed mean =357, N= The number of scores in the distribution=50)

g. Apply the formula for calculation of standard deviation

$$= \text{Standard Deviation} = (\sigma) = i / N [\sqrt{N \cdot \Sigma fx'^2 - (\Sigma fx')^2}]$$

$$= 5/50 [\sqrt{50 \times 357 - (-17)^2}]$$

$$= 5/50 [\sqrt{17850 - 289}]$$

$$=5/50 \sqrt{17561}$$

$$=1/10 \times 132.5179$$

Standard Deviation= (σ) = 13.2517

(b)

Table 6.58 The Data Grouped into a Frequency Distribution

S.No.	Class Intervals	Mid-Point(x)	(f)	(x')	(fx')	(fx' ²)
1.	170-175	172	1	5	5	25
2.	165-170	167	1	4	4	16
3.	160-165	162	2	3	6	18
4.	155-160	157	3	2	6	12
5.	150-155	152	4	1	4 (+25)	4
6.	145-150	147	7	0	0	0
7.	140-145	142	2	-1	-2	2
8.	135-140	137	2	-2	-4	8
9.	130-135	132	2	-3	-6	18
10.	125-130	127	3	-4	-12	48
11.	120-125	122	2	-5	-10	50
12.	115-120	117	1	-6	-6 (-40)	36
			N=30		$\Sigma fx' = -15$	$\Sigma fx'^2 = 237$

The formula used for calculating variance for grouped set of data with the help of assumed mean deviation scores is:

$$\text{Standard Deviation} = (\sigma) = i / N [\sqrt{N \cdot \Sigma fx'^2 - (\Sigma fx')^2}]$$

(i= Width of the class-interval, x' = Deviation of the raw measures/scores from the assumed mean, N= The number of scores in the distribution)

a. We take the set of grouped set of data in Table 6.58:

b. Calculate deviation of raw data from the assumed mean= x'

c. Calculate the sum of deviation of raw data from the assumed mean is =fx'=-15

c. Calculate the sum of squared deviation of raw data from the assumed mean =
fx'²= 237

$$\text{Standard Deviation} = (\Sigma) = i / N [\sqrt{N \cdot \Sigma fx'^2 - (\Sigma fx')^2}]$$

$$\Sigma fx' = -15 \quad \Sigma fx'^2 = 237$$

(i= Width of the class-interval=5, x' = Deviation of the raw measures/scores from the assumed mean, sum of squared deviation of raw data from the assumed mean =237, N= The number of scores in the distribution=30)

g. Apply the formula for calculation of standard deviation

$$\begin{aligned} \text{Standard Deviation} = (\sigma) &= i / N [\sqrt{N \cdot \Sigma fx'^2 - (\Sigma fx')^2}] \\ &= 5/30 [\sqrt{30 \times 237 - (-15)^2}] \end{aligned}$$

$$\begin{aligned}
&= 5/30[\sqrt{7110-225}] \\
&= 5/30 \sqrt{6885} \\
&= 1/6 \times 82.9759
\end{aligned}$$

Standard Deviation= (σ) = 13.8293

(c)

Table 6.59 The data grouped into a frequency distribution

S.No.	Class Intervals	Mid-Point(x)	(f)	(x')	(fx')	(fx' ²)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.	165-170	167	2	5	10	50
2.	160-165	162	2	4	8	32
3.	155-160	157	2	3	6	18
4.	150-155	152	3	2	6	12
5.	145-150	147	3	1	3(+33)	3
6.	140-145	142	6	0	0	0
7.	135-140	137	2	-1	-2	2
8.	130-135	132	1	-2	-2	4
9.	125-130	127	2	-3	-6	18
10.	120-125	122	3	-4	-12	48
11.	115-120	117	2	-5	-10	50
12.	110-115	112	2	-6	-12 (-44)	72
			N=30		$\Sigma fx' = -11$	$\Sigma fx'^2 = 309$

The formula used for calculating Standard Deviation for grouped set of data with the help of assumed mean deviation scores is:

$$\text{Standard Deviation} = (\sigma) = i / N [\sqrt{N \cdot \Sigma fx'^2 - (\Sigma fx')^2}]$$

(i= Width of the class-interval, x' = Deviation of the raw measures/scores from the assumed mean, N= The number of scores in the distribution)

a. We take the set of grouped set of data in Table 6.59:

b. Calculate deviation of raw data from the assumed mean= x'

c. Calculate the sum of deviation of raw data from the assumed mean is =fx'=-11

c. Calculate the sum of squared deviation of raw data from the assumed mean = fx'²= 309

d. Apply the formula for calculation of

$$\text{Standard Deviation} = (\sigma) = i / N [\sqrt{N \cdot \Sigma fx'^2 - (\Sigma fx')^2}]$$

$\Sigma fx' = -11$	$\Sigma fx'^2 = 309$
--------------------	----------------------

(i= Width of the class-interval=5, x' = Deviation of the raw measures/scores from the assumed mean, sum of squared deviation of raw data from the assumed mean =309, N= The number of scores in the distribution=30)

g. Apply the formula for calculation of standard deviation

$$\begin{aligned}
 \text{Standard Deviation} &= (\sigma) = \frac{1}{N} [\sqrt{N \sum fx'^2 - (\sum fx')^2}] \\
 &= 5/30 [\sqrt{30 \times 309 - (-11)^2}] \\
 &= 5/30 [\sqrt{9270 - 121}] \\
 &= 5/30 \sqrt{9149} \\
 &= 1/6 \times 95.6504
 \end{aligned}$$

$$\text{Standard Deviation} = (\sigma) = 15.9416$$

let us check our progress-4

(i) Calculate the Pearson's r from the following ungrouped data.

(a) Table 6.60 Two Variables of Ungroup Data

X	55	64	62	69	72	56	65	59	60	64	57	58
Y	52	60	65	56	69	51	56	58	55	58	59	59

When size of sample is small and data are not grouped then Pearson's r may be calculated as follows:

$$r_{xy} = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}}$$

Here, x= deviations of X measures from assumed mean

y= deviations of Y measures from assumed mean

We may calculate the Pearson's product moment 'r' from the following data for the two variables for 12 nursing students of our college.

Calculation:

X	Y	x= X- AM	y = Y- AM	x ²	y ²	Xy
55	52	55-62= -7	52-58= -6	49	36	42
64	60	64-62= 2	60-58= 2	4	4	4
62 (AM)	65	62-62= 0	65-58= 7	0	49	0
69	56	69-62= 7	56-58= -2	49	4	-14
72	69	72-62= 10	69-58= 11	100	121	110
56	51	56-62= -6	51-58= -7	36	49	42
65	56	65-62= 3	56-58= -2	9	4	-6
59	58(AM)	59-62= -3	58-58= 0	9	0	0
60	55	60-62= -2	55-58= -3	4	9	6
64	58	64-62= 2	58-58= 0	4	0	0
57	59	57-62= -5	59-58= 1	25	1	-5
58	59	58-62= -4	58-58= 0	16	0	0
		Σx = -3	Σy = 1	Σx² = 305	Σy² = 277	Σxy = 179

$$\begin{aligned}
 r_{xy} &= \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}} \\
 &= \frac{12 \times 179 - (-3) \times (1)}{\sqrt{[(12 \times 305) - (-3)^2][12 \times 277 - (1)^2]}}
 \end{aligned}$$

$$\begin{aligned}
&= 2148 - (-3) / \sqrt{[3660 - 9][3324 - 1]} \\
&= 2151 / \sqrt{[3651][3323]} \\
&= 2151 / \sqrt{12132273} \\
&= 2151 / 3483.14125 \\
&= 0.6175
\end{aligned}$$

(b) Table 6.61 Two Variables of Ungroup Data

X	56	64	62	69	72	56	65	59	60	64	59	58
Y	53	60	65	56	69	51	56	58	55	58	60	59

When size of sample is small and data are not grouped then Pearson's r may be calculated as follows:

$$r_{xy} = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}}$$

Here, x= deviations of X measures from assumed mean

y= deviations of Y measures from assumed mean

We may calculate the Pearson's product moment 'r' from the following data for the two variables for 12 nursing students of our college.

Calculation:

X	Y	x= X- AM	y = Y- AM	x ²	y ²	XY
56	53	56-62= -6	53-58= -5	36	25	30
64	60	64-62= 2	60-58= 2	4	4	4
62 (AM)	65	62-62= 0	65-58= 7	0	49	0
69	56	69-62= 7	56-58= -2	49	4	-14
72	69	72-62= 10	69-58= 11	100	121	110
56	51	56-62= -6	51-58= -7	36	49	42
65	56	65-62= 3	56-58= -2	9	4	-6
59	58(AM)	59-62= -3	58-58= 0	9	0	0
60	55	60-62= -2	55-58= -3	4	9	6
64	58	64-62= 2	58-58= 0	4	0	0
59	60	59-62=-3	60-58=2	9	4	-6
58	59	58-62=-4	59-58=1	16	1	-4
		$\sum x = 0$	$\sum y = 4$	$\sum x^2 = 276$	$\sum y^2 = 270$	$\sum xy = 162$

$$\begin{aligned}
r_{xy} &= \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}} \\
&= \frac{12 \times 162 - (0) \times (4)}{\sqrt{[(12 \times 276) - (0)^2][12 \times 270 - (4)^2]}} \\
&= \frac{1944 - (0)}{\sqrt{[3312 - 0][3240 - 16]}} \\
&= \frac{1944}{\sqrt{[3312][3224]}} \\
&= \frac{1944}{\sqrt{10677888}} \\
&= \frac{1944}{3267.7037} \\
&= 0.5949
\end{aligned}$$

(c) Table 6.62 Two Variables of Ungroup Data

X	56	64	62	69	72	56	65	59	60	64	59
Y	53	60	65	56	69	51	56	58	55	58	60

When size of sample is small and data are not grouped then Pearson's r may be calculated as follows:

$$r_{xy} = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}}$$

Here, x= deviations of X measures from assumed mean

y= deviations of Y measures from assumed mean

We may calculate the Pearson's product moment 'r' from the following data for the two variables for 11 nursing students of our college.

Calculation:

X	Y	x= X- AM	y = Y- AM	x ²	y ²	Xy
56	53	56-62= -6	53-58= -5	36	25	30
64	60	64-62= 2	60-58= 2	4	4	4
62 (AM)	65	62-62= 0	65-58= 7	0	49	0
69	56	69-62= 7	56-58= -2	49	4	-14
72	69	72-62= 10	69-58= 11	100	121	110
56	51	56-62= -6	51-58= -7	36	49	42
65	56	65-62= 3	56-58= -2	9	4	-6
59	58(AM)	59-62= -3	58-58= 0	9	0	0
60	55	60-62= -2	55-58= -3	4	9	6
64	58	64-62= 2	58-58= 0	4	0	0
59	60	59-62=-3	60-58=2	9	4	-6
		Σx =4	Σy =3	Σx² =260	Σy² =269	Σxy=166

$$\begin{aligned}
 r_{xy} &= \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}} \\
 &= \frac{11 \times 166 - (4) \times (3)}{\sqrt{[(11 \times 260) - (4)^2][11 \times 269 - (3)^2]}} \\
 &= \frac{1826 - (12)}{\sqrt{[2860 - 16][2959 - 9]}} \\
 &= \frac{1814}{\sqrt{[2844][2950]}} \\
 &= \frac{1814}{\sqrt{8389800}} \\
 &= \frac{1814}{2896.5151} \\
 &= 0.6262
 \end{aligned}$$

(ii) Calculate the Pearson's r from the following grouped data.

(a) Calculation of Pearson's r from grouped data

Table 6.63 Marks Obtained in Psychology and Sociology

Marks obtained in Psychology (Y)	Marks obtained in Sociology (X)					
	25-35	35-45	45-55	55-65	65-75	Total
25-35	8	6	-	-	-	14
35-45	5	15	12	-	-	32
45-55	-	10	15	7	-	32
55-65	6	-	5	10	4	25
65-75	-	4	-	-	5	9
Total	19	35	32	17	9	112

Process of calculation of Pearson's r from grouped data

Following steps may be taken for calculation of Pearson's r from grouped data:

- (i) First of all, we take the step deviations of x variable and denote these deviations by 'dx'.
- (ii) After it, we take the step deviations of y variable and denote these deviations by 'dy'.
- (iii) Multiply dx, dy and the respective frequency of each cell and write the figure obtained in the left-hand corner of each cell.
- (iv) Now we add together all the values as calculated in step (iii) and obtain the total $\Sigma f dx dy$.
- (v) Multiply the frequencies of x variable by dx and obtain the total $\Sigma f dx$.
- (vi) Multiply f dx by dx and obtain the total $\Sigma f dx^2$.
- (vii) Multiply the frequencies of y variable by dy and obtain the total $\Sigma f dy$.
- (viii) Multiply f dy by dy and obtain the total $\Sigma f dy^2$.
- (ix) Substitute the values of $\Sigma f dx dy$, $\Sigma f dx$, $\Sigma f dx^2$, $\Sigma f dy$, $\Sigma f dy^2$ and N in the following formula.
- (x) As a result, we obtain the value of 'r'.

Formula:
$$r = \frac{\Sigma f dx dy \cdot N - (\Sigma f dx \cdot \Sigma f dy)}{[\{\sqrt{\Sigma f dx^2 \cdot N - (\Sigma f dx)^2}\} \{\sqrt{\Sigma f dy^2 \cdot N - (\Sigma f dy)^2}\}]}$$

Example: Calculate coefficient of correlation between marks obtained in Anatomy and Physiology of nursing students of a college of nursing in mid-term examination.

Calculation:

Marks obtained in Anatomy (Y)	Marks obtained in Physiology (X)										
	25-35	35-45	45-55	55-65	65-75	f	fdy	fd ² y	fdxdy		
	dx	-2	-1	0	+1	+2					
	dy										
25-35	-2	4	8	6	12	32	2				
35-45	-1	2	5	1	15	10	0				
45-55	0	0	0	0	0	0	0				
55-65	+1	-2	6	-12	0	10	0				
65-75	+2	-2	4	-8	0	20	0				
f	19	35	32	17	9	112(N)	(-)17Σfdy	149 Σfd ² y	87 Σfdxdy		
fdx	-38	-35	0	17	18	(-)38Σfdx					
fd ² x	76	35	0	17	36	164 Σfd ² x					
fdxdy	30	19	0	10	28	87 Σfdxdy					

$$\begin{aligned}
r &= \{\Sigma f dx dy . N - (\Sigma f dx . \Sigma f dy)\} / [\sqrt{\{\Sigma d^2 x . N - (\Sigma f dx)^2\}} \sqrt{\{\Sigma f d^2 y . N - (\Sigma f dy)^2\}}] \\
&= \{87 \times 112 - (-38) \times -17\} / [\sqrt{\{164 \times 112 - (-38)^2\}} \sqrt{\{149 \times 112 - (-17)^2\}}] \\
&= 9744 - (646) / \sqrt{[18368 - 1444]} \sqrt{[16688 - 289]} \\
&= 9098 / \sqrt{16924 \times 16399} \\
&= 9098 / \sqrt{277536676} \\
&= 9098 / 16659.4320 \\
&= 0.5461
\end{aligned}$$

(b) Calculation of Pearson's r from grouped data

Table 6.64 Marks Obtained in Microbiology and Nutrition

Marks obtained in Microbiology (Y)	Marks obtained in Nutrition (X)					
	30-40	40-50	50-60	60-70	70-80	Total
30-40	7	3	-	-	-	10
40-50	4	17	13	-	-	34
50-60	-	12	16	7	-	35
60-70	-	-	8	11	2	21
70-80	-	-	-	5	1	6
Total	11	32	37	23	3	106

Process of calculation of Pearson's r from grouped data

Following steps may be taken for calculation of Pearson's r from grouped data:

- First of all, we take the step deviations of x variable and denote these deviations by 'dx'.
- After it, we take the step deviations of y variable and denote these deviations by 'dy'.
- Multiply dx, dy and the respective frequency of each cell and write the figure obtained in the left-hand corner of each cell.
- Now we add together all the values as calculated in step (iii) and obtain the total $\Sigma f dx dy$.
- Multiply the frequencies of x variable by dx and obtain the total $\Sigma f dx$.
- Multiply fdx by dx and obtain the total $\Sigma f d^2 x$.
- Multiply the frequencies of y variable by dy and obtain the total $\Sigma f dy$.
- Multiply fdy by dy and obtain the total $\Sigma f d^2 y$.
- Substitute the values of $\Sigma f dx dy$, $\Sigma f dx$, $\Sigma f d^2 x$, $\Sigma f dy$, $\Sigma f d^2 y$ and N in the following formula.
- As a result, we obtain the value of 'r'.

Formula: $\{\Sigma f dx dy . N - (\Sigma f dx . \Sigma f dy)\} / \sqrt{\{\Sigma f d^2 x . N - (\Sigma f dx)^2\}} \sqrt{\{\Sigma f d^2 y . N - (\Sigma f dy)^2\}}$

Example: Calculate coefficient of correlation between marks obtained in Anatomy and Physiology of nursing students of a college of nursing in mid-term examination.

Solution:

Marks obtained in Anatomy (Y)	Marks obtained in Physiology (X)										
	30-40	40-50	50-60	60-70	70-80	f	fdy	fd ² y	fdxdy		
	dx										
	-2	-1	0	+1	+2						
30-40	4	2				10	-20	40	34		
	7	3	-	-	-						
	28	6									
40-50	2	1	0			34	-34	34	25		
	4	17	13	-	-						
	8	17	0								
50-60		0	0	0		35	0	0	0		
	-	12	16	7	-						
		0	0	0	0						
60-70			0	1	2	21	21	21	15		
	-	-	8	11	2						
			0	0	4						
70-80				2	4	6	12	24	14		
	-	-	-	5	1						
				10	4						
f	11	32	37	23	3	106(N)	(-)21Σfdy	119 Σfd ² y	88 Σfdxdy		
fdx						(-)25Σfdx					
fd ² x	-22	-32	0	23	6	111 Σfd ² x					
fdxdy	44	32	0	23	12	88 Σfdxdy					
	36	23	0	21	8						

$$\begin{aligned}
r &= \frac{\sum f dx dy \cdot N - (\sum f dx \cdot \sum f dy)}{\sqrt{\{\sum f d^2 x \cdot N - (\sum f dx)^2\} \{\sum f d^2 y \cdot N - (\sum f dy)^2\}}} \\
&= \frac{\{88 \times 106 - (-25)(-21)\}}{\sqrt{[111 \times 106 - (-25)^2][119 \times 106 - (-21)^2]}} \\
&= \frac{(9328 - 525)}{\sqrt{(11766 - 625)(12614 - 441)}} \\
&= \frac{8803}{\sqrt{11141 \times 12173}} \\
&= \frac{8803}{\sqrt{135619393}} \\
&= \frac{8803}{11645.5739} \\
&= 0.75590
\end{aligned}$$

c. Table 6.62 Marks Obtained in Pathology and Biochemistry

Marks obtained in Pathology (Y)	Marks obtained in Biochemistry (X)					
	25-35	35-45	45-55	55-65	65-75	Total
25-35	8	4	3	-	-	15
35-45	5	18	12	-	-	35
45-55	-	13	15	9	-	37
55-65	-	-	9	12	3	24
65-75	-	-	-	7	2	9
Total	13	35	39	28	5	120

Process of calculation of Pearson's r from grouped data

Following steps may be taken for calculation of Pearson's r from grouped data:

- (i) First of all, we take the step deviations of x variable and denote these deviations by 'dx'.
- (ii) After it, we take the step deviations of y variable and denote these deviations by 'dy'.
- (iii) Multiply dx, dy and the respective frequency of each cell and write the figure obtained in the left-hand corner of each cell.
- (iv) Now we add together all the values as calculated in step (iii) and obtain the total $\sum f dx dy$.
- (v) Multiply the frequencies of x variable by dx and obtain the total $\sum f dx$.
- (vi) Multiply fdx by dx and obtain the total $\sum f d^2 x$.
- (vii) Multiply the frequencies of y variable by dy and obtain the total $\sum f dy$.
- (viii) Multiply fdy by dy and obtain the total $\sum f d^2 y$.
- (ix) Substitute the values of $\sum f dx dy$, $\sum f dx$, $\sum f d^2 x$, $\sum f dy$, $\sum f d^2 y$ and N in the following formula.
- (x) As a result, we obtain the value of 'r'.

$$\text{Formula: } \frac{\{\sum f dx dy \cdot N - (\sum f dx \cdot \sum f dy)\}}{\sqrt{\{\sum f d^2 x \cdot N - (\sum f dx)^2\} \{\sum f d^2 y \cdot N - (\sum f dy)^2\}}}$$

Example: Calculate coefficient of correlation between marks obtained in Anatomy and Physiology of nursing students of a college of nursing in mid-term examination.
Calculation:

Marks obtained in Anatomy (Y)	Marks obtained in Physiology (X)									
	25-35	35-45	45-55	55-65	65-75	f	fdy	fd ² y	fdxdy	
	dx	-2	-1	0	+1	+2				
	dy									
25-35	-2	4	2	0						
		8	4	3	-	15	-30	60	40	
		32	8	0						
35-45	-1	2	1	0						
		5	18	12	-	35	-35	35	28	
		10	18	0						
45-55	0		0	0						
		-	13	15	9	37	0	0	0	
			0	0	0					
55-65	+1			0	1					
		-	-	9	12	24	24	24	18	
				0	12					
65-75	+2				2					
		-	-	-	7	9	18	36	22	
					14					
f		13	35	39	28	120(N)	(-)23Σfdy	155 Σfd ² y	108Σfdxdy	
fdx		-26	-35	0	28	(-)23Σfdx				
fd ² x		52	35	0	28	135 Σfd ² x				
fdxdy		42	26	0	26	108Σfdxdy				

$$\begin{aligned}
 r &= \{ \sum fdxdy.N - (\sum fdx.\sum fdy) \} / \sqrt{ \{ \sum fd^2x.N - (\sum fdx)^2 \} \{ \sum fd^2y.N - (\sum fdy)^2 \} } \\
 &= \{ 108 \times 120 - (-23)(-23) \} / \sqrt{ [135 \times 120 - (-23)^2] [155 \times 120 - (-23)^2] } \\
 &= (12960 - 529) / \sqrt{ [(16200 - 529)] [(18600 - 529)] } \\
 &= 12431 / \sqrt{(15671) \times (18071)} \\
 &= 12431 / \sqrt{283190641} \\
 &= 12431 / 16828.2691029113 \\
 &= 0.7386
 \end{aligned}$$

(iii) Calculate co-efficient of correlation by ranking method from following data.

(a) In a class test of two subjects, B.Sc. Nursing students, obtained marks in both subjects as given in following table:

Table 6.66 Marks Obtained in Psychology and Sociology

Code of Students	Marks obtain in Psychology (X)	Marks obtain in Sociology (Y)	Difference (D)	D ²
A	3	4	-1	1
B	5	6	-1	1
C	6	5	1	1
D	1	10	-9	81
E	9	8	1	1
F	2	3	-1	1
G	4	2	2	4
H	10	9	1	1
I	7	1	6	36
J	5	7	-2	4
K	7	8	-1	1
L	9	7	2	4
M	7	8	-1	1
N	8	6	2	4
O	6	8	-2	4
				$\Sigma D^2 = 145$

The Calculation of Rank order correlation

$$p(\rho) = 1 - 6\Sigma D^2 / N(N^2 - 1)$$

(Where, ΣD^2 = Sum of squared differences between ranks = 145, N = Number of paired ranks = 15)

$$\begin{aligned}
 p(\rho) &= 1 - 6 \times 145 / 15(15^2 - 1) \\
 &= 1 - 870 / 15(225 - 1) \\
 &= 1 - 870 / 15(224)
 \end{aligned}$$

$$= 1 - 870/3360$$

$$= (3360-870)/3360$$

$$= 2490/3360$$

$$= 0.7410$$

(b) In a class test of two subjects, B.Sc. Nursing students, obtained marks in both subjects as given in following table:

Table 6.67 Marks Obtained in Psychology and Sociology

Code of Students	Marks obtain in Psychology (X)	Marks obtain in Sociology (Y)	Difference (D)	D ²
A	3	4	-1	1
B	5	6	-1	1
C	6	5	1	1
D	1	10	-9	81
E	9	8	1	1
F	2	3	-1	1
G	4	2	2	4
H	10	9	1	1
I	7	1	6	36
J	5	7	-2	4
K	7	8	-1	1
L	9	7	2	4
M	7	8	-1	1
				$\Sigma D^2 = 137$

The Calculation of Rank order correlation

$$p(\rho) = 1 - 6\Sigma D^2 / N(N^2 - 1)$$

(Where, ΣD^2 = Sum of squared differences between ranks = 137, N = Number of paired ranks = 13)

$$p(\rho) = 1 - 6 \times 137 / 13(13^2 - 1)$$

$$= 1 - 822 / 13(169 - 1)$$

$$= 1 - 822 / 2184$$

$$= (2184 - 822) / 2184$$

$$= 1362 / 2184$$

$$= 0.6236$$

(c) In a class test of two subjects, B.Sc. Nursing students, obtained marks in both subjects as given in following table:

Table 6.68 Marks Obtained in Psychology and Sociology

Code of Students	Marks obtain in Psychology (X)	Marks obtain in Sociology (Y)	Difference (D)	D ²
A	3	4	-1	1
B	5	6	-1	1
C	6	5	1	1
D	1	10	-9	81
E	9	8	1	1
F	2	3	-1	1
G	4	2	2	4
H	10	9	1	1
I	7	1	6	36
J	5	7	-2	4
K	7	8	-1	1
L	9	7	2	4
M	7	8	-1	1
N	5	9	-4	16
O	9	7	2	4
P	6	8	2	4
Q	7	9	2	4
				$\Sigma D^2 = 165$

The Calculation of Rank order correlation

$$p(\rho) = 1 - 6\Sigma D^2 / N(N^2 - 1)$$

(Where, ΣD^2 = Sum of squared differences between ranks = 165, N = Number of paired ranks = 17)

$$\begin{aligned}
 p(\rho) &= 1 - 6 \times 165 / 17(17^2 - 1) \\
 &= 1 - 990 / 17(289 - 1) \\
 &= 1 - 990 / 17(288) \\
 &= 1 - 990 / 4896 \\
 &= (4896 - 990) / 4896 \\
 &= 3906 / 4896 \\
 &= 0.7977
 \end{aligned}$$

BIBLIOGRAPHY

1. Chowdhury Sunanda S Ray. (2010). Text Book of Nursing Research.1st Edition, Kumar Publishing House, New Delhi
2. Garret Henry E. (2005). Statistics in Psychology and Education. 12th Edition, Paragon International Publishers, New Delhi. ISBN:81-89253-00-X
3. Gupta SL, GuptaHitesh Gupta. (2011). Research Methodology (Text and Cases with SPSS Application), 2nd Edition, International book House Pvt. Ltd. New Delhi. ISBN:978-81-910642-5-4
4. Khandelwal RS, Agrawal Raju, Juniwal Ravi, Ahmad Tauseff, Sharma Esha. (2010). Research Method in Management, 1st Edition, Ajmera Book Company, Jaipur. ISBN:978-93-80722-00-9
5. Kothari CR, Garg Gaurav. (2019). Research methodology: Methods and techniques. New Age International Publishers.New Delhi. ISBN:978-93-86649-22-5
6. Kumar Ranjit. (2005). Research Methodology (A Step-By-Step Guide for Beginners). 2nd Edition, Pearson education, New Delhi. ISBN:978-81-317-0496-7
7. Oswal ML, Agrawal NP, Modi SS, Bhargava PK, Tiwari Ajay, Meena Udai Singh. (2019-20). Business Statistics, 1st Edition, RBD Publishing House, Jaipur. ISBN:978-81-8142-080-2
8. Polit Denise F. (2010). Nursing Research: Generating and Assessing Evidence for Nursing Practice.9th Edition, Wolters Kluwer India Pvt. Ltd., New Delhi.ISBN:13-978-81-8473-533-8
9. Sharma Suresh K. (2014). Nursing Research& Statistics. Reed Elsevier India Private Limited. New Delhi ISBN: 978-81-312-3308-5
10. Singh Arun Kumar. (2006). Tests, Measurements and Research Methods in Behavioural Sciences. 5th Edition, Bharti Bhawan Publishers & Distributors, New Delhi. ISBN:81-7709-743-1

Notes

Notes

Notes