Indira Gandhi
National Open University
School of Social Work

MSW-006
Social Work Research

Block

4

DATA PROCESSING AND ANALYSIS

EXPERT COMMITTEE

Prof. Surendra Singh (Late)
Former Vice-Chancellor
Kashi Vidyapeeth, Varanasi

Prof. Thomas Kalam
St. John's Medical College
Bangalore

Dr. Mukul Srivastava
Dr. B. R. Ambedkar
University, Agra

Prof. Jyoti Kakkar
Jamia Millia Islamia
New Delhi

Prof. Gracious Thomas
IGNOU, New Delhi

Prof. Sanjai Bhatt
Delhi University
Delhi

Dr. Joseph Xavier
Indian Social Institute
Bangalore

Dr. Usha John
Loyala College
Trivandrum

Prof. Ranjana Sehgal
Indore School of Social
Work, Indore

Prof. Anjali Gandhi
Jamia Millia Islamia
New Delhi

Dr. Leena Mehta
M.S. University
Vadodara

Prof. Archana Dassi
Jamia Millia Islamia
New Delhi

Dr. Beena Antony
Delhi University
Delhi

EXPERT COMMITTEE (Revision)

Prof. Gracious Thomas
School of Social Work
IGNOU, Delhi

Dr. D.K.Lal Das
R.M College
Hyderabad

Prof. P.K.Ghosh
Department of Social Work
Visva Bharti University,
Shantiniketan

Prof. C.P.Singh
Department of Social Work
Kurukshetra University

Mr. Joselyn Lobo
Roshni Nilaya
Mangalore

Prof. Ranjana Sehgal
Indore School of Social Work
Indore

Dr. Asiya Nasreen
Department of Social Work
Jamia Millia Islamia University,
Delhi

Dr. Bishnu Mohan Dash
B.R.Ambedkar College
Delhi University

Dr. Rose Nembiakkim
School of Social Work
IGNOU, Delhi

Dr. Saumya
School of Social Work
IGNOU, Delhi

Dr. G.Mahesh
School of Social Work
IGNOU, Delhi

Dr. N.Ramya
School of Social Work
IGNOU, Delhi

COURSE PREPARATION TEAM

Unit Writers
Dr. Monika Jauhari
and adapted from
ES-315 and MRD-004

Course Editor
Prof. J.S. Gandhi
J.N.U. New Delhi

Block Editor &
Programme Coordinator
Prof. Gracious Thomas
IGNOU, New Delhi

COURSE PREPARATION TEAM (Revision)

Unit Writers
Dr. Monika Jauhari
and adapted from
ES-315 and MRD-004

Course Editor
Prof. Jyoti Kakkar
Jamia Millia Islamia
New Delhi

Programme Coordinator
Dr. Saumya

PRINT PRODUCTION

Mr. Kulwant Singh
Assistant Registrar (Publication)
SOSW, IGNOU

February, 2019 (Revised Edition)

ISBN-978-81-266-3499-6

Further information, about the Indira Gandhi National Open University courses may be obtained from the University's office at Maidan Garhi, New Delhi-110 068.

Printed and published on behalf of the Indira Gandhi National Open University by Director, School of Social Work

Laser Typeset by : Rajshree Computers, V-166A, Bhagwati Vihar, (Near Sec. 2, Dwarka), Uttam Nagar, New Delhi-110059

BLOCK INTRODUCTION

"Data Processing and Analysis" is the fourth block of the course on Social Work Research. This block has four units and each unit provides you relevant information about processing, analysis of data, descriptive and inferential statistics and reporting in research.

Unit 1 First unit of this block is on **'Data Processing and Analysis'.** In this unit we have discussed about the processing of quantitative data, coding of data and preparing a master chart as well as analysis of quantitative data.

Unit 2 'Descriptive Statistics' is the second unit of this block. The purpose of this unit is to provide a detailed discussion on measures of central tendency, mean, median, mode, quartile deviation, mean deviation standard and deviation which are essential components of statistics.

Unit 3 is on **'Inferential Statistics'.** This unit will provide you information on measures of relationship, product moment, correlation, coefficient of correlation, chi-square, regression analysis and measures of differences. We have also discussed T-test, paired, sample, independent samples, F-test, and testing of hypothesis.

Unit 4 is the last unit in this block titled: **'Reporting of Research'.** This unit will provide you knowledge about what, why and how of reporting a research. In this unit we have discussed about how to begin and write the research work, its main body, tables, figures, bibliography, references and appendices of the research.

The entire block on data processing and analysis is very informative and essential in meaningfully carrying out a research in Social Work.

4th Blank

UNIT 1 DATA PROCESSING AND ANALYSIS

Structure

1.0 OBJECTIVES

After the completion of this Unit, you should be able to;

- classify, categorize, and re-categorize data;
- prepare code book and master chart;
- prepare analytic model for data analysis; and
- prepare various types of tables for presenting data.

1.1 INTRODUCTION

In previous unit, we discussed about methods and tools of data collection. Once data are collected the researcher turns his focus of attention on its processing. In this chapter, we have discussed about one of the most important stages of the research process, i.e. data- processing and analysis.

Processing of Quantitative Data

Introduction

Once data are collected, the researcher turns his/her focus of attention on its processing. A researcher has to make his/her plan for each and every stage of the research process. As such, a good researcher makes a perfect plan of processing and analysis of data. To some researchers, data processing and analysis is not a very serious activity. They feel, many a time, that data processing is a job of a computer assistant. Consequently, they remain contented with the results given by computer assistant which may not help them to achieve their objectives. To avoid such situations, it is essential that data processing is planned in advance and assistant instructed accordingly. In this chapter, we will discuss about data processing and data analysis.

Data processing refers to certain operations such as editing, coding, computing of the scores, preparation of master charts, etc.

Editing of Data

After collection of filled-in questionnaires/interview schedules, editing of entries therein are not only necessary but also useful in making subsequent steps simpler. Many a times, a researcher or the assistant either misses entries in the questionnaires or enters wrong responses. This sort of discrepancies can be resolved by editing the schedule meticulously. Another problem comes up at the time of tabulation of data when researcher asks for tabulation of responses from consecutive questions. In cases where data are not edited there is bound to be inconsistency in the tabulation. In the process of editing, the researcher has to be very careful about consecutive questions having 'not applicable' as a response.

Check Your Progress I

Note: Use the space given below for your answers.

1) Describe the importance of editing of data.

.....

.....

.....

.....

.....

.....

.....

.....

.....

1.3 CODING OF DATA

Coding of data involves assigning of number to each response of the question. The purpose of giving numbers is to translate raw data into numerical data, which may be counted and tabulated. The task of researcher is to give numbers to responses carefully. The coding scheme will vary according to the type of questions. For example, a close- ended question may be already coded and hence it has to be just included in the code book whereas coding of open-ended questions involves operations such as classification of major responses and developing a response category of 'others' for responses which were not given frequently. The classification of responses is primarily based on similarities or differences among the responses. Usually in the case of open-ended questions, to classify responses, researcher looks for major characteristics of the responses and puts them accordingly. In case of attitude scales, researcher has to keep in mind the direction or weightage of responses. For example, a response 'strongly agree' is coded as '5'. The subsequent codes would be in order. Therefore, if there are responses like 'agree', 'undecided', 'disagree' and 'strongly disagree' they have to be coded as 4, 3, 2, and 1. Alternatively, if strongly agree is coded as –2, the subsequent responses would be coded as –1, zero, +1 and +4. The matrix questions have to be coded taking into consideration each cell as one variable. For example, if the column of

matrix represents employment status, namely 'permanent' and 'temporary' and row represents employers or type of employer, namely government and private, the first cell would represent a variable 'government- permanent'. The second cell would represent 'government-temporary' and so on. In order to demonstrate the points discussed above a section of a code book used in a study is reproduced in Table given below.

Table 1.1 : Code Book

Q.No.	Variable No.	Variable Label	Responses	Code
2	V1	Age	Actual	-
3	V2	Designation	Worker Supervisor Manager	1 2 3
4	V3	Establishment	Public Private	1 2
5	V4	Level of Education	Graduate Intermediate High School Middle School Primary Illiterate Other	1 2 3 4 5 6 7
6	V5	Marital Status	Married Unmarried Widow Divorce	1 2 3 4
36	V35	Nature of work	Yes No	1 2
	V36	Duration of work	Yes No	1 2
	V37	Wages	Yes No	1 2
	V38	Promotion	Yes No	1 2
43	V42	Attitude of Employer Preference for male employees	Agree Undecided Disagree	1 2 3

1.4 PREPARING A MASTER CHART

After a code book is prepared, the data can be transferred either to a master chart or directly to computer through a statistical package. Going through master chart to computer is much more advantageous than entering data directly to computers because one can check the wrong entries in the computer by comparing 'data listing' as a computer output and master chart. Entering data directly to

computer is disadvantageous as there is no way to check wrong entries which will show inconsistencies in tabulated data at the later stages of tabulation. A sample of master chart prepared in accordance with the code book is presented in Table given.

Table 1.2 : Master Chart

Respondent Number	Variable Lables									
	Age	Designation	Establishment	Level of Education	Marital Status	Nature of Work	Duration of Work	Wages	Promotions	Attitude of Employer
					Variable Number					
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	23	3	4	2	2	3	4	1220	3	4
2	28	2	3	4	1	5	5	1547	3	3
3	23	4	4	4	2	1	4	1922	1	3
4	28	1	3	3	1	2	5	1847	2	5
5	23	3	4	2	2	1	7	2922	1	1
6	24	3	1	1	3	1	1	1220	1	1
7	22	3	1	1	3	1	2	1547	1	1
8	24	2	1	1	3	1	2	1922	2	1
9	45	2	2	1	3	1	2	1847	2	2
10	35	3	1	2	2	2	1	2922	3	3
11	36	1	2	2	2	3	2	1847	4	4
12	34	2	3	1	1	2	3	2922	4	3
13	35	2	4	2	2	2	4	1220	3	2
14	46	1	3	3	1	1	4	1547	3	1
15	45	2	3	3	1	2	5	1922	2	1
16	23	1	4	3	2	3	5	1847	1	1

Analysis of Quantitative Data

The purpose of data analysis is to prepare data as a model where relationships between the variables can be studied. Analysis of data is made with reference to objectives of the study and research questions, if any. It is also designed to test the hypothesis. Analysis of data involves re-categorisation of variables, tabulation, explanation and casual inferences.

The first step in data analysis is a critical examination of the processed data in the

form of frequency distribution. This analysis is made with a view to draw meaningful and precise inferences and generalizations.

The process of categorization in accordance with the objectives and hypothesis of the study is arrived at with the help of frequency distributions. Re-categorisation is a process to arrange categories with the help of statistical techniques. The process may involve addition of new category or merging existing categories. This helps researcher to justify the tabulation. We have seen earlier that responses to a statement may be assigned scores or weightage. These scores or weightage are summated and re-categorised on the bases of certain ranks like high, medium and low. The basic principle in the process of categorisation or re-categorisation is that categories obtained must be exhaustive and mutually exclusive. In other words, the categories have to be independent without overlapping.

Tabulation

Tabulation is a process of presenting data in a tabular form in such a way as to facilitate comparisons and reflect relations. In other words, it is an arrangement of data in rows and columns. This also helps the researcher to perform statistical operation on the data to draw inferences. Tabulation can be generally in the form of univariate, bivariate, tri or multivariate tables. Accordingly, analysis proceeds in the form of univariate analysis, bivariate analysis and tri or multivariate analysis.

Setting Up the Analytic Model

Before we begin the analysis of data we have to look once again at the objectives of the research study and set up analytic models. These models are diagrammatic presentation of variables and their interrelationships.

Let us hypothesise that 'Life Satisfaction is likely to be associated with Socio-economic Status of elderly'.

The two variables in the hypothesis are 'Life Satisfaction' and the 'Socio-economic Status'. The relation between the two variables can be diagrammatically presented as follows:

Socio-economic Status → **Life Satisfaction**

Further, it is hypothesised that the bi-variate relationships between the variables are affected by another variable 'social support'. Let us suppose that 'social support' has been categorised into three, namely, high, medium and low. This can be described as follows :

Social Support: High

Socio-economic Status — Life Satisfaction

Social Support: Medium

Socio-economic Status — Life Satisfaction

Social Support : Low

Socio-economic Status — Life Satisfaction

With the analytic model described above the researcher can proceed to analyse the data as discussed in the following sections :

Univariate Analysis

Univariate analysis refers to tables which hold data relating to one variable. Univariate tables which are more commonly known as frequency distribution tables show how frequently an item is repeated occured. Examples of frequency tables are given below. The distribution may be symmetrical or asymmetrical. The characteristics of the sample while examining the percentages, further properties of a distribution can be found out by various measures of central tendencies. However, researcher is required to decide which is most suited for the analysis. To know how much is the variation, the researcher has to calculate measures of dispersion. (The details about all these measures are given in unit 2 of this block.)

Usually, frequency distribution tables are prepared to examine each of the independent and dependent variables. Tables given below present two independent variables and one dependent variable.

Table 1.3 : Socio-Economic Status of Elderly (Independent Variable)

Socio-economic Status	Distribution of Respondents	
	Frequencies	Percentage
High	110	39.3
Medium	106	37.9
Low	64	21.8
Total	280	100.0

Table 1.4 : Social Support (Independent Variable)

Social Support	Distribution of Respondents	
	Frequencies	Percentage
High	142	57.7
Medium	86	30.7
Low	52	14.6
Total	280	100.0

Table 1.5 : Life Satisfaction of Elderly (Dependent Variable)

Life Satisfaction	Distribution of Respondents	
	Frequencies	Percentage
High	78	27.9
Medium	134	47.9
Low	68	24.2
Total	280	100.0

A frequency distribution of a single variable is the frequency of observation in each category of a variable. For example, an examination of the pattern of response to variable 'socio-economic status' in Table 1.3 would provide a description of the number of respondents who belong to 'high', 'medium' and 'low' socio-economic status. Let us consider the frequency distribution (Tables 1.3, 1.4 and 1.5) which describes the 'socio-economic status', 'social support' and 'life satisfaction' of respondents. The tables have four rows, the first three being the categories of variables, which appear in the left-hand columns and the right hand columns show the number of observation in each category. The last rows are the totals of all frequencies appearing in tables. To analyse the data, it is necessary to convert the frequencies into percentages that can be interpreted meaningfully.

Bivariate Analysis

As a researcher you might be interested in knowing the relationships between the variables. To know the relationship, the data pertaining to the variables are cross-tabulated. Hence, a bivariate table is also known as cross table. A bivariate table presents data of two variables with column percentages or row percentages. An example of a bivariate table is given below:

Table 1.6: Levels of Socio-economic Status and Life Satisfaction of the Respondents

Socio-economic	Life satisfaction			Total
	High	Medium	Low	
High	94 (66.2)	9 (10.5)	7 (13.5)	110 (39.3)
Medium	37 (26.1)	58 (67.4)	11 (21.2)	106 (37.9)
Low	11 (7.7)	19 (22.1)	34 (65.3)	64 (22.8)
Total	142 (100.0)	86 (100.0)	52 (100.0)	280 (100.0)

The table presents data with regard to two variables namely 'socio-economic status' and the level of 'life satisfaction'. First row presents data with regard to respondents who were from high Socio-economic Status. The second row presents data about who were from medium socio-economic status . Similarly, the first column gives data pertaining to respondents who preferred high life satisfaction. The second column presents data of respondents who preferred medium life satisfaction and the last column represents the respondents who preferred low life satisfaction. For example, the first cell (in the left-hand corner) represents 94 respondents who were from high Socio-economic Status and high life satisfaction.

The association between two variables can be explained either by comparing the percentages of respondents column wise or row wise. The relationship between the variables can also be examined by various statistical techniques (The details about all these measures are given in chapter 10.) depending upon the level of measurement of the data.

Check Your Progress II

Note: Use the space given below for your answers.

1) Describe the purpose of bivariate analysis of data.

.....

.....

.....

.....

.....

.....

.....

Trivariate Analysis

Sometimes researcher might be interested in knowing whether there is a third variable which is affecting the relationships between two variables. In such cases the researcher has to examine the bivariate relationship by controlling the effects of third/variable**.** This is performed in two ways. One way of controlling the effects of a third/variable is to prepare partial tables and examine the bi-variate relationship. The second method of assessing the effects of a third/variable is to compare the co-efficient of partial correlations. Let us take an example. In the above table, if researcher wants to examine whether there is effect of 'social supports' on the bivariate relationship, he may prepare three partial tables giving data relating to socio-economic status and life satisfaction for high, medium and low 'social supports'.

Table 1.7 : Levels of Socio-economic Status and Life Satisfaction of the Respondents

Social Support = High (N = 142)

Socio-economic Status	Life satisfaction			Total
	High	Medium	Low	
High	7 (21.9)	10 (21.8)	3 (9.3)	20
Medium	13 (40.6)	26 (33.3)	9 (28.1)	48
Low	12 (37.5)	42 (53.8)	20 (62.5)	74
Total	32	78	32	142

Table 1.8 : Levels of Socio-economic Status and Life Satisfaction of the Respondents

Social Support = Medium(N = 86)

Socio-economic Status	Life satisfaction			Total
	High	Medium	Low	
High	8 (36.4)	11 (25.6)	4 (19.0)	23
Medium	9 (40.9)	17 (39.5)	8 (38.1)	34
Low	5 (34.9)	15 (34.9)	9 (42.9)	29
Total	22	43	21	86

Table 1.9 : Levels of Socio-economic Status and Life Satisfaction of the Respondents

Social Support = Low (N = 52)

Socio-economic Status	Life satisfaction			Total
	High	Medium	Low	
High	14 (58.3)	7 (53.8)	7 (46.7)	28
Medium	8 (33.3)	4 (30.8)	6 (40.0)	18
Low	2 (8.3)	2 (15.4)	2 (13.3)	6
Total	24	13	15	52

On examination of these three partial tables, if the researcher finds that bivariate relationships don't hold good he/she may infer that it is the third variable i.e., social supports which is affecting the bivariate relationship. In the partial tables for higher social support, the proportion of people perceiving high life satisfaction are those who are having high socio economic status. The similar trend can be noticed in the remaining two partial tables. This means that social support does not affect the bivariate relationships between socio economic status and life satisfaction.

In the other method, the researcher tries to find out the effect of third variable on the bivariate relationship with the help of partial correlation. To find out whether the third variable affects the bivariate relationship between socio economic status (X) and life satisfaction (Y), he/she considers social support of respondents.

To examine this, he/she calculates the correlation coefficients between the 'socio economic status' and 'life satisfaction' keeping 'social support' as a constant. The results are presented below.

X_1 = socio-economic status (independent variable)

X_2 = life satisfaction' (dependent variable)

X_3 = social support (control variable)

$\bullet_{1.2}$ (Correlation coefficient between X_1 and X_2) = .70

$\bullet_{1.3}$ (Correlation coefficient between X_1 and X_3) = .40

$\bullet_{2.3}$ (Correlation coefficient between X_2 and X_3) = .50

The researcher on examination of the results of the partial correlation finds that even after controlling the third variable, the relation between the first two variables i.e. socio-economic status and life satisfaction is statistically significant. Thus, he/she concludes that the relationship between socio-economic status and life satisfaction is not spurious.

Multivariate Analysis

When a researcher is interested in assessing the joint effect of three or more variables he/she uses the techniques of multivariate analysis. The most common statistical technique used for multivariate analysis is regression analysis (For this statistical technique reader is advised to refer books listed at the end of this Unit). In the first step of multivariable analysis, the researcher has to obtain the correlation between the variables which are having statistically significant correlation. These variables are put in the regression analysis. One important point in applying correlation and regression analysis is that the data must be measured on ratio or interval level. Another point a researcher has to keep in mind is that these two statistical techniques are based on certain assumptions. Hence, before applying these techniques, the researcher has to see whether the sample selected by him fulfills the prerequisite conditions.

Check Your Progress III

Note: Use the space given below for your answers.

1) What is tri-variate analysis. Discuss the process of examining the effect of third variable on the bivariate relationship.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

1.6 LET US SUM UP

In this unit, we discussed various stages of processing and analysis of quantitative data. The main points are as follows:

1) Data processing refers to certain operations such as editing, coding, computing of the scores, preparation of master charts, etc.

2) Many a times, a researcher or the assistants either miss entries in the questionnaires or enter responses, which are not legible. This sort of discrepancies can be resolved by editing the questionnaires.

3) Coding of data involves assigning of numbers to each response of the question. The purpose of giving numbers is to translate raw data into numerical data, which may be counted and tabulated.

4) With the help of a code book, the data can be transferred to a master chart.

5) Tabulation is a process of presenting data in rows and columns in such a way as to facilitate comparisons and show the involved relations.

6) Keeping in view the objectives of research study, we set up analytic models which diagrammatically present variables and their interrelationships.

7) Univariate analysis refers to tables which give data relating to one variable.

8) A bivariate table presents data of two variables in column percentages and row percentages simultaneously.

9) Trivariate analysis is undertaken to know whether there is a third variable which is affecting the relationships between two variables.

10) When a researcher is interested in assessing the joint effect of three or more variables he uses the techniques of multivariate analysis.

1.7 FURTHER READINGS AND REFERENCES

Bailey, Kenneth, D. (1978), *Methods of Social Research,* The Free Press, London.

Baker, L. Therese (1988), *Doing Social Research,* McGraw Hill, New York.

Blalock, Hubert, M. (1972), *Social Statistics,* Mc Graw Hill, New York.

Doby J.T. (et. al.) (1953), *An Introduction to Social Research,* Harrisburg: Stackpole Co., London.

Kerlinger Fred R. (1964), *Foundations of Behavioural Research,* Surjeet Publications, Delhi.

Lal Das, D.K. (2000), *Practice of Social Research: A Social Work Perspective,* Rawat Publications, Jaipur.

Monete, Duane R. (et. al.) (1986), *Applied Social Research,* Tool For the Human Services, Holt, Chicago.

Nachmias, D and Nachmias, C. (1981), *Research Methods in the Social Sciences,* St. Martins Press, New York.

Rafael, J., Engel. and Russell, K., Schutt. (2013). The Practice fo Research in Social Work - Third Edition. Sage Publication Inc. USA.

Sellitz. G. et. al. (1973), *Research Methods in Social Relations,* Rinehart and Winston (3rd edition) *Holt, New York.*

Sjoberg, G. and Nett., B. (1968), *A Methodology of Social Research,* Harper & Row, New York.

UNIT 2 DESCRIPTIVE STATISTICS

Structure

2.0 OBJECTIVES

After the completion of this Unit, you should be able to:

- compute measures of central tendency; and
- compute measures of dispersion.

2.1 INTRODUCTION

In Unit 1, you learnt about the data processing and data analysis. The aim of this chapter is to introduce various descriptive statistical methods used in data analysis.

If you have a set of data, what will be your next step? You would like to understand that data. Statistics is all about gaining an understanding of data. And what one does to gain understanding – one operates on that data by means of average and graphs. If you look beyond average and graphs, you will see that data are actually numbers in some context. Therefore, you need to do something extra to understand the context.

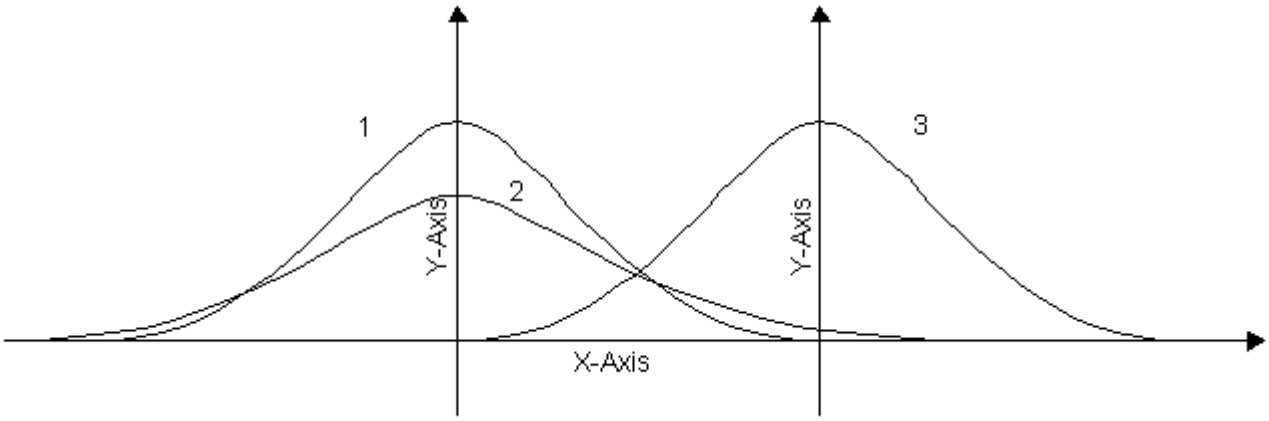
Knowledge of statistics helps managers in two ways. First, the knowledge allows the managers to be able to analyse the data as per his research objectives and draw inferences, which in turn, enable him to expand and improve not only the knowledge base of his profession but also enables him to make his practice effective. Second, as a consumer of research it enables him to understand the analysis of data and appreciate the statistical procedures used in research reports.

Statistical procedures have increasingly become a part of the manager's support system. These procedures are used to enhance effectiveness and efficiency of the services offered by professionals. One of the significant reasons for this is the availability of various statistical packages which are generally quite easy to perform. Consequently, in the recent years, there has been a sizeable increase in the use of microcomputers by business managers in analysing data about their clients/beneficiaries.

2.2 MEASURES OF CENTRAL TENDENCY

It is often essential to represent a set of data by means of a single number which, in its way, is descriptive of the entire set. Obviously, the figure which is used to represent a whole series should neither have the lowest value in the series nor the highest value, but a value somewhere between these two limits, possibly in the

centre. Such figures are called measures of central tendency or average. In other words, numerical description of a distribution begins with a measure of its center or average.



Comparison of central tendency of curves

In the above figure you can see three curves. The center of curve 3 differs from centers of curve 1 and 2. But although curves 1 and 2 have different shapes their centers are the same. Why is this so? We will explain the first statement here and leave the second statement for section 2.3.

There are three important measures of central tendency used in social work, the mean or arithmetic mean, the median and the mode.

Arithmetic Mean

The arithmetic mean is by far the most common of all the averages. It is relatively easy to calculate, simple to understand and is widely used in social work research. The arithmetic mean is defined as the ratio of the sum of the values of all the items and the total number of items.

In calculating arithmetic mean of a continuous series, we take the mid-value of each class as representative of that class (and it is presumed that the frequencies of that class are concentrated on mid-point), multiply the various mid-values by their corresponding frequencies and sum of the products is divided by sum of the frequencies. The following illustration will explain the method:

Table 2.1 : Distribution of Street Children by their Daily Earnings

S.No.	Daily Earnings (in Rs.)	Number of Street Children
1	110-130	15
2	130-150	30
3	150-170	60
4	170-190	95
5	190-210	82
6	210-230	75
7	230-250	23
	Total	**380**

Table 2.2

Daily Earnings (in Rs.)	Mid-values (m)	Number of Street Children (f)	Deviation from Assumed Mean 180 (d_x = m–180)	Step Deviation ($d=d_x/20$)	Total Deviation (fd)
110-130	120	15	–60	–3	–45
130-150	140	30	–40	–2	–60
150-170	160	60	–20	–1	–60
170-190	180	95	0	0	0
190-210	200	82	+20	+1	+82
210-230	220	75	+40	+2	+150
230-250	240	23	+60	+3	+69
		N = 380			**Σfd=136**

Assumed Mean = 180

Class Interval (i)= 130 – 110 = 20

Arithmetic Mean

Where 'a' stands for the assumed mean, $\sum fd$ for the sum of total deviations, N for total number of frequencies and 'i' for class interval. Now substituting the values in the formula from the table we get :

$$\bar{X} = 180 + \frac{136}{380} \times 20 = 180 + 7.16$$

Arithmetic Mean

Arithmetic Mean = Rs.187.16 (approximately)

Median

The median is another simple measure of central tendency. Sometimes you would like to locate the position of the middle item when data have been arranged or sometimes you would like to divide a group of students into quartiles (the ones that divide the series into 4 parts) by locating the individuals who have exactly 25 per cent of the class below them, exactly 75 per cent below them etc. These measures are known as positional averages. The median is perhaps the most important of such positional averages. We define the median as the size of the middle item when the items are arranged in ascending or descending order of magnitude. This means that median divides the series in such a manner that there are as many items above or larger than the middle one as there are below or smaller than it.

In continuous series, we do not know every observation. Instead, we have record of the frequencies with which the observations appear in each of the class-intervals as in the following Table. Nevertheless, we can compute the median by determining which class-interval contains the median.

Table2.3 : Daily Earning of Street Children

Daily Earnings (in Rs.)	Number of Street Children (f)	Cumulative frequencies (CF)
110 – 130	15	15
130 – 150	30	45
150 –170	60	105
170 – 190	95	200
190 – 210	82	282
210 – 230	75	357
230 – 250	23	380
	N = 380	

In the case of data given in Table the median value is that value on either side of which N/2 or 380/2 or 190th item lie. Now the problem is to find the class interval containing the 190th item. The cumulative frequency for the first three classes is only 105. But when we move to the fourth class interval 95 items are added to 105 for total of 200. Therefore, the 190th item must be located in this fourth class-interval (the interval from Rs.170 – Rs.190). So the *median class* is the class whose cumulative frequency just exceeds N/2.

The median class (Rs.170 – Rs.190) for the series contains 95 items. For the purpose of determining the point, which has 190 items on each side, we assume that these 95 items are evenly spaced over the entire class interval 170-190. Therefore, we can interpolate and find the values for 190th item. First, we determine that the 190th item is the 95th item in the median class:

$$190 - 105 = 85$$

Then we can calculate the width of the 95 equal steps from Rs.170 to Rs.190 as follows:

$$\frac{190-170}{95} = 0.21053 \text{ (Approximately)}$$

The value of 85th item is 0.2105 × 85 = 17.89. If this (17.89) is added to the lower limit of the median class, we get 170 + 17.89 = 187.89. This is the median of the series.

This can be put in the form of formula:

Where M_e = median,

L = lower limit of the class in which median lies

N = total number of items

PCF= cumulative frequency of the class prior to the median class.

f = frequency of the median class

i = class interval of the median class.

Table 2.4 : Daily Wages of Workers

Daily Earnings (in Rs.)	Number of Street Children (f)	Cumulative frequencies (CF)
110 – 130	15	15
130 – 150	30	45
150 –170	60	105 = PCF
170 – 190 (L = 170)	95 = f	200
190 – 210	82	282
210 – 230	75	357
230 – 250	23	380
	N = 380	

$$M_e = L + \frac{\frac{N}{2} - PCF}{f} \times i$$

$$= 170 + \frac{190 - 105}{95} (190 - 170)$$

Me

$$Me = 170 + \frac{85}{95} \times 20$$

= 170 + (0.8947 × 20) = 187-89 (approximately)

Mode

Another measure, which is sometimes used to describe the central tendency of a set of data, is the mode. It is defined as the value that is repeated most often in the data set. In the following series of values: 71, 73, 74, 75, 75, 75, 78, 78, 80 and 82, the mode is 75, because 75 occur more often than any other value (three times). In grouped data, the mode is located in the class where the frequency is greatest. The mode is more useful when there are a larger number of cases and when data have been grouped.

Calculation of Mode

Calculation of mode involves two steps. First, the process of grouping locates the class of maximum concentration with the help of grouping method. The procedure is as follows:

i) First the frequencies are added in two's in two ways: (a) by adding frequencies of item numbers 1 and 2; 3 and 4; 5 and 6 and so on, and (b) by adding frequencies of item numbers 2 and 3; 4 and 5; 6 and 7 and so on.

ii) Then the frequencies are added in three's. This can be done in three ways: (a) by adding frequencies of item numbers 1, 2 and 3; 4,5 and 6; 7,8 and 9; and so on. (b) by adding frequencies of item numbers 2,3 and 4; 5, 6 and 7; 8,9 and 10; and so on and (c) by adding frequencies of item numbers 3,4 and 5; 6,7 and 8; 9,10 and 11 and so on.

If necessary, grouping of frequencies can be done in four's and five's also. After grouping, the size of items containing maximum frequencies is circled. The item value, which will contain the maximum frequency the largest number of times, is the mode of the series. This is shown in Tables given below.

After this the value of mode is interpolated by the use of the following formula.

Where M_0 stands for the mode, L is the lower limit of the modal class (the class that has the maximum frequency), f_0 stands for the frequencies of the class preceding modal class, f_1 stands for the frequencies of the modal class, f_2 for the frequencies of the class succeeding modal class and i stand for the class interval of the modal class.

Illustration: Determine mode of the following distribution.

Table 2.5 : Monthly Expenditure

Monthly Expenditure (In Rs)	Number of Families
100 – 200	5
200 – 300	6 = f_0
300 – 400 L = 300	15 = f_1
400 – 500	10 = f_2
500 – 600	5
600 – 700	4
700 – 800	3
800 – 900	2
Total	N = 50

Table 2.6 : Location of Modal Class by Grouping

Monthly Expenditure (In Rs)	f(1)	(2)	(3)	(4)	(5)	(6)
100 – 200	5	11		26		
200 – 300	6		21		31	
300 – 400	15	25				30
400 – 500	10		15	19		
500 – 600	5	9			12	
600 –700	4		7			9
700 –800	3	5				
800 –900	2					

Table 2.7 : Analysis Table

Column	**Class Containing Maximum Frequency**							
	100-200	200-300	300-400	400-500	500-600	600-700	700-800	800-900
1			1					
2			1	1				
3		1	1					
4	1	1	1					
5		1	1	1				
6			1	1	1			
No.of times a class occurs	1	3	6	3	1			

$M_0 = 300 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times 100$

Therefore, 330-400 group is the modal group. Using the formula of interpolation, viz.,

$= 300 + 900/14 = 300 + 64.29$

$= 364.29$ (approximately)

Check Your Progress I

Note: Use the space given below for your answers.

1) Illustrate the methods of expressing the class intervals with the help of examples.

2) Name and define the three measures of central tendency.

3) Compute (i) Mean (ii) Median and (iii) Mode for the following frequency distribution.

Class Interval	f
195-199	1
190-194	2
185-189	4
180-184	5
175-179	8
170-174	10
165-169	6
160-164	4
155-159	4
150-154	2
145-149	3
140-144	1

.....

.....

.....

.....

.....

.....

.....

.....

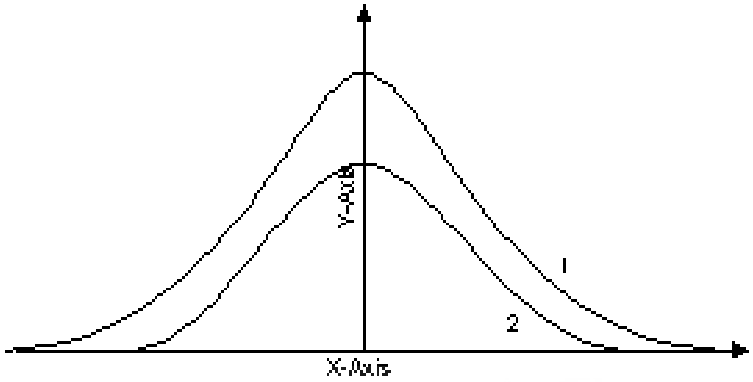
.....

2.3 MEASURES OF DISPERSION

Introduction

In research, one often wishes to know the extent of homogeneity and heterogeneity among respondents with respect to a given characteristic. If you consider a set of data whose values are heterogeneous, such set of data is characterized by the

heterogeneity of values. In fact, the extent to which they are heterogeneous or vary among themselves is of great importance in statistics. Measures of central tendency describe one important characteristic of a set of data typically but they do not tell us anything about this other basic characteristic. Consequently, we need ways of measuring heterogeneity – the extent to which data are dispersed and the measures, which provide this description, are called measures of dispersion or variability.



Comparison of Dispersion of two curves

The picture above is a representation of the "spread" of the data. Curve 1 has a wider spread than curve 2. In other words, data in curve 1 is more scattered and hence the measure of location is less representative of the data than the one in curve 2. This is the answer to the second statement that we discussed in section 2.2

The important measures of dispersion are range, quartile deviation, mean deviation and standard deviation.

Range

The range is defined as the difference between the highest and lowest values. Mathematically,

$$R(\text{Range}) = m_H - m_L$$

Where m_H and m_L stand for the highest and the lowest value respectively. Thus, for the data set; 10, 22, 20, 14 and 14 the range would be the difference between 22 and 10, i.e., 12. In case of grouped data, we take the range as the difference between the midpoints of the extreme classes. Thus, if the midpoint of the lowest interval is 150 and that of the highest are 850 the range will be 700.

The only advantage of range, the measure that is seldom used, is that it can be easily calculated and readily understood. Despite this advantage, it is generally not a very useful measure of dispersion. Its main drawback is that it does not tell us anything about the dispersion of values, which are intermediate between the two extremes.

Quartile Deviation or Semi-Inter-Quartile Range

Another measure of dispersion is the semi-inter-quartile range, commonly known as quartile deviation. Quartiles are the points which divide the array or series of values into four equal parts, each of which contains 25 per cent of the items in the distribution. The quartiles are then the highest values in each of these four parts. Inter-quartile range is the difference between the values of first and the third quartiles.

Thus, semi-inter-quartile range or quartile deviation is calculated by the following formula

$$Q.D. = \frac{Q_3 - Q_1}{2}$$

Where Q_1 and Q_3 stand for first and the third quartiles.

Calculation of Quartile Deviation (QD)

Table 2.8 : Monthly Expenditure

Monthly Expenditure (In Rs.)	Number of Families
100 – 200	5
200 – 300	6
300 – 400	15
400 – 500	10
500 – 600	5
600 – 700	4
700 – 800	3
800 – 900	2
Total	**N = 50**

Table 2.9

Sl. No.	Expenditure (2)	Number of Families (3)	Cumulative frequency(CF)
1	100 – 200	5	5
2	200 – 300	6	11 = c_1
3	$Q_1 \rightarrow$ 300 – 400	15 = f_1	26
4	400 – 500	10	36 = c_3
5	$Q_3 \rightarrow$500 – 600	5 = f_3	41
6	600 – 700	4	45
7	700 – 800	3	48
8	800 – 900	2	50
	Total	**N = 50**	

For Q_1, Let us calculate N/4 = 50/4 = 12.5, so Q_1 will fall in the class 300–400

$$Q_1 = L + \frac{N/4 - c_1}{f_1} \times i$$

Q_1

$$Q_1 = 300 + \frac{1.5}{15} \times 100$$

$$= 300 + (0.1 \times 100)$$

$$= 300 + 10 = 310$$

Similarly for Q_3, Let us calculate 3N/4 = 150/4 = 37.5, so Q_3 will fall in the class 500–600

$$Q_3 = L + \frac{3N/4 - c_3}{f_3} \times i$$

Q_3

$$Q_3 = 500 + \frac{1.5}{5} \times 100$$

$$= 500 + (0.3 \times 100) = 500 + 30 = 530$$

$$Q_3 - Q_1 = 530 - 310 = 220$$

$$= 500 + \frac{37.5 - B}{15} \times 100 \qquad QD = \frac{Q_3 - Q_1}{2}$$

$$QD = \frac{220}{2} = 110$$

Quartile Deviation is an absolute measure of dispersion. If quartile deviation is to be used for comparing the dispersion of series it is necessary to convert the absolute measure to a coefficient of quartile deviation.

Symbolically, coefficient of Quartile Deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1/2} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$

Applying this to the preceding illustration we get, Coefficient of Quartile Deviation,

$$\text{Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{530 - 310}{530 + 310} = \frac{220}{840} = 0.26 \text{ (approximately)}$$

Mean Deviation

Range and quartile deviation suffer from a serious drawback; they are calculated by taking into consideration only two values of a series. These two measures of dispersion are not based on all observations of the series. As a result, the composition of the series is entirely ignored. To avoid this complication, dispersion is calculated taking into consideration all the observations of the series in relation

to a central value. The method of calculating dispersion is called the method of averaging deviations (Mean Deviation). As the name suggests, it is the arithmetic average of the deviations of various items from a measure of central tendency.

Calculation of Mean Deviation

Table 2.10 : Monthly Expenditure

Monthly Expenditure (In Rs.)	Number of Families
100 – 200	5
200 – 300	5
300 – 400	15
400 – 500	10
500 – 600	5
600 – 700	4
700 – 800	3
800 – 900	3
Total	N = 50

Table 2.11 : Monthly Expenditure

	Monthly Expenditure (In Rs)	Mid Value	Number of (f)	Comulative frequency Families	Deviation from median 400 |d|	f |d|
	100-200	150	5	5	250	1250
	200-300	250	5	10=C	150	750
Median	300-400	350	15=f	25	50	750
Group	400-500	450	10	35	50	500
	500-600	550	5	40	150	750
	600-700	650	4	44	250	1000
	700-800	750	3	47	350	1050
	800-900	850	3	50	450	1350
			N = 50			**7400**

Step I : Calculate the median of the distribution:

$$M_e = L + \frac{\frac{N}{2} - C}{F} \times i$$

M_e

$$M_e = 300 + \frac{25-10}{15} \times 100$$

$$M_e = 300 + \frac{15}{15} \times 100$$

$= 300 + (1 \times 100) = 300 + 100 = 400$

Median = 400

Step II : Find mid-points of each class

$$Midpoint = \frac{100+200}{2} = \frac{300}{2} = 150, ...$$

Step III : Find absolute deviation |d | of each midpoints from median (400)

|d| = | 150 – 400 | = 250, ...

Find total absolute deviation by multiplying the frequency of each class by the deviation of its mid – points from the median (f |d|) = 5 × 250 = 1250, . .

Find the sum of products of frequency and deviations $\Sigma(f|d|) = 7400$

Compute Mean Deviation= λ (X)

$$\lambda(X) = \frac{\sum f|d|}{N} = \frac{7400}{50} = 148$$

Coefficient of Mean Deviation

To compare the mean deviation of series the coefficient of mean deviation or relative mean deviation is calculated by dividing the mean deviation by that measure of central tendency about which deviations were calculated. Thus,

Coefficient of Mean Deviation = $\frac{\lambda(X)}{M_e}$

Applying this formula to the previous example, we have,

Coefficient of Mean Deviation =

Standard Deviation

The most useful and frequently used measure of dispersion is standard deviation or root-mean square deviation about the mean. The standard deviation is defined as the square root of the arithmetic mean of the squares of the deviations about the mean.

Symbolically.

$$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \times i$$

Where 'i' stands for the common factor or the magnitude of the class-interval. The following example would illustrate this formula;

Table 2.12 : Monthly Expenditure

S.No.	Monthly Expenditure (In Rs)	Number of Families(f)
1	100 – 200	5
2	200 – 300	6
3	300 – 400	15
4	400 – 500	10
5	500 – 600	5
6	600 – 700	4
7	700 – 800	3
8	800 – 900	2
		N = 50

Table 2.13

S. No.	Monthly Expenditure (In Rs.)	Mid values (m)	Number of Families (f)	Step deviation from assumed Average 450 (d)	fd	d^2	fd^2
1	100 – 200	150	5	-3	-15	9	45
2	200 – 300	250	6	-2	-12	4	24
3	300 – 400	350	15	-1	-15	1	15
4	400 –500	450	10	0	0	0	0
5	500 –600	550	5	+1	5	1	5
6	600 –700	650	4	+2	8	4	16
7	700 –800	750	3	+3	9	9	27
8	800 - 900	850	2	+4	8	16	32
			N = 50		Σfd = –12		Σfd^2 = 164

Step	Procedure	Application to Table 13
1	Find the mid-points of the various classes	$Midpoint = \frac{100+200}{2} = \frac{300}{2}$ = 150, ...
2	Assume a mid-point as average, preferably at the centre	450 = assumed average.
3	Take the difference of each mid-point from the assumed average (450) and divide them by the magnitude of the class interval to get step deviation (d)	(1) 150 – 450 = –300/100 = –3…
4	The deviations are multiplied by the frequency of each class (fd)	(–3) (5) = –15 (–2) (6) = –12…
5	Find the aggregate of products of step 4 (Σfd)	Σfd = –12
6	Square the deviations (d^2)	(–3) (–3) = 9, . . .
7	Squared deviations are multiplied bythe respective frequencies (fd^2)	9 x 5 = 45, . . .
8	Find the aggregate of products of step 7 (Σfd)	$\Sigma fd^2 = 164$
9	Compute standard deviation with the help of the formula $\sigma = \sqrt{3.28 - 0.0576} \times 100$	$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \times i$ $\sigma = \sqrt{\frac{164}{50} - \left(\frac{-12}{50}\right)^2} \times 100$ $\sigma = \sqrt{3.2224} \times 100$ =1.795 ×100 = 179.51 (approximately)

Coefficient of Variation

The standard deviation represents measure of absolute dispersion. It is also necessary to measure the relative dispersion of two or more distributions. When the standard deviation is related to its mean, it measures relative dispersion.

Karl Pearson has worked out a simple measure of relative dispersion, which is generally known as the coefficient of variation.

$$C.V.\ or\ V = \frac{\sigma}{\bar{X}} \times 100$$

The mean & coefficient of variation for the problem in Table 2.13 are:

$$\overline{X} = 450 + \frac{-12}{50} \times 100$$

$$= 450 + (-12) \times 2 = 450 - 24 = 426$$

$$C.V.\ or\ V = \frac{\sigma}{\overline{X}} \times 100 = \frac{179.51}{426} \times 100$$

$$= 0.42138 \times 100 = 42.138 \text{ per cent (approximately)}$$

Check Your Progress II

Note: Use the space given below for your answers.

1) Name and define the three measures of variability (range, variance, and standard deviation)

2) Compute standard deviation for the following frequency distribution

Class Interval	f
195-199	1
190-194	2
185-189	4
180-184	5
175-179	8
170-174	10
165-169	6
160-164	4
155-159	4
150-154	2
145-149	3
140-144	1

.....

.....

.....

.....

.....

.....

.....

2.4 LET US SUM UP

1) Measures of (i) central tendency, (ii) variability, and (iii) relationship are the three types of descriptive statistical measures.

2) Mean, median and mode are the three measures of central tendency.

3) Mean is the arithmetic average of a distribution. It is obtained by dividing the sum of all values of observation by the total number of values.

4) Median is a point in an array, above and below which one half of the values or measures fall. If the values are ungrouped and their number is small, the values are arranged in order of magnitude and the middle value is determined by counting up half, the value of N. When the number of values is odd, the mid-value is the median. When the number of values is even, the median is the mid-point between the two middle values.

5) Mode is the most frequently occurring value in a distribution. If only one value occurs a maximum number of times, the distribution is said to have one mode (uni-modal). A two-mode distribution is bi-modal, and more than a two-mode distribution is called multi-modal.

   In a simple ungrouped series of measures or values, the crude mode is that single measure or value which occurs most frequently.

6) The range is the difference between the two extreme values or measures in a distribution.

2.5 FURTHER READINGS AND REFERENCES

Alston, Margaret. and Bowles, Wendy. Research for Social Workers – an introduction to methods. Rawat Publications. (New Delhi & Jaipur) (2003).

Baker, L. Therese. *Doing Social Research,* New York: McGraw Hill, 1988.

Black, Allen & Babbie E. *Research Methodology for Social Work*, Belmont, California, Wadsworth, 1989.

Blalock, H.M.*Social Statistics*, Auckland. Mc Graw Hill, 1988.

Elhance, D.N. *Fundamental of Statistics*, Allahabad, Kitab Mahal, 1984.

Festinger. L. and Katz. D. (eds.) *Research Methods in the Behavioural Sciences,* New York : The Dryden Press,1953.

Freud, J.E. *Modern Elementary Statistics*, New Delhi, Prentice Hall, 1977.

Garret, H.E. *Statistic in Physical & Education*, Bomb: Vakils, 1966.

Gupta, S.P. *Statistical Methods*, New Delhi, S.Chand, 1980.

Kidder, Lauise H. *Research Methods in Social Relations,* New York: Holt, 1981.

Krishef, Curtis H. *Fundamentals Statistics for Human Services and Social Work*, Boston: Duxbury Press, 1987.

Lal Das, D.K. (2000), *Practice of Social Research: A Social Work Perspective,* Rawat Publications, Jaipur.

Levin, Richard I. *Statistics for Management* Delhi : Pearson Prentice Hall, 2005.

Monette, Duana R. et. al. *Applied Social Research: Tool For the Human Services,* Chicago: Holt, 1986.

Nachmias D and Nachmias C. *Research Methods in the Social Sciences*, New York: St. Martins press, 1981.

Ostle, B. and Mensing R.W. *Statistics in Research*, New Delhi, Oxford, 1975.

Philip, A.E. et. al. *Social Work Research*, and The Analysis of Social Data, Oxford: Peragon Press, 1975.

Sanders, D.H. et. al. *Statistics A Fresh Approach*, New Delhi: Mc Graw Hill, 1975.

Thyer, Bruce. The handbook of SocialWork Research Methods (2010). Sage Publications, Inc. (USA)

UNIT 3 INFERENTIAL STATISTICS

Structure

3.0 OBJECTIVES

After the completion of this Unit, you should be able to:

- compute measures of relationship; and
- formulate a hypothesis and apply tst ofsignificance to arrive at a decision.

3.1 INTRODUCTION

In unit 2 you learnt about descriptive statistics. The aim of this unit is to ponder upon various inferential statistical methods. Till now you have learnt about univariate distributions (involving one variable only). In practice, you may often come across certain series where you may be interested in measuring two characteristics e.g. educational level and intelligence quotient of staff of a consultancy firm. Such type of distribution is called a bivariate distribution. In such situations one is interested in whether there is any relation between the two variables (Correlation). The measure of the average relationship between two variables is Regression.

Another important aspect here is the study of tests of significance, which enables us to decide about the population parameters on the basis of observed sample results. Unit 3 discusses these aforesaid methods.

3.2 MEASURES OF RELATIONSHIP

So far, you have learnt about various statistical methods concerned with the description and analysis of a single variable. In research, you often wish to know the relationship among two or more variables in the data with one another. One may wish to know, for instance, whether the expenditure on research and development (R&D) is somehow related to the companies' profits. In other words, one may be interested in ascertaining whether a change in one variable is associated with change in another variable. For example, whether the higher expenditure R & D is associated with higher profits; whether the salary of executives is correlated with their work performance. The most frequently used measure of estimating association among variables is the coefficient of correlation (r).

The correlation coefficient (r) to be discussed here was introduced by Karl Pearson and is often referred to as product – moment correlation. Coefficient of correlation is calculated to identify the extent or degree of correlation between two variables.

Karl Pearson's Coefficient of Correlation

The formula to compute the degree and direction of correlation of Karl Pearson's is as under;

$$r = \frac{\frac{1}{N}\sum d_x d_y - \left(\frac{\sum d_x}{N}\right)\left(\frac{\sum d_y}{N}\right)}{\sqrt{\frac{\sum d_x^2}{N} - \left(\frac{\sum d_x}{N}\right)^2} \times \sqrt{\frac{\sum d_y^2}{N} - \left(\frac{\sum d_y}{N}\right)^2}}$$

Where d_x is the deviation of various items of the first variable from assumed average and d_y the corresponding deviations of the second variable from assumed average and N stands for the number of pairs of items. We explain the application of the formula in the following example.

Table 3.1 : Marks Obtained in Research and Dissertation

Sl.No	Research	Dissertation
1	40	42
2	26	29
3	70	66
4	32	24
5	52	58
6	30	28
7	48	46
8	48	50
9	42	44
10	62	66

Table 3.2 : Calculation of Karl Pearson's Coefficient of Correlation

Sl.	Research	Deviations from ass. Average (d_x=x–48)	dx^2	Dissertation	Deviations from ass. Average (d_y=y–46)	d_y^2	$d_x.d_y$
1	40	-8	64	42	-4	16	+32
2	26	-22	484	29	-17	289	+374
3	70	+22	484	66	+20	400	+440
4	32	-16	256	24	-22	484	+352
5	52	+4	16	58	+12	144	+48
6	30	-18	324	28	-18	324	+324
7	48	0	0	46	0	0	0
8	48	0	0	50	+4	16	0
9	42	-6	36	44	-2	4	+12
10	62	+14	196	66	+20	400	+280
N=10	$\bar{X}$ =45	Σd_x = –30	Σd_x^2= 1860	$\bar{Y}$= 45.3	Σd_y= –7	Σd_y^2= 2077	$\Sigma d_x d_y$ 1862

Table 3.3

Step	Procedure	Application to Table 2
1	Assume average of variable X and Y. (Preferably in the neighborhood of X)	Ass. Ave (X) = 48 Ass. Ave (Y) = 46
2	Find deviations of item values of X and Y from assumed averages (d_x) and (d_y)	e.g. for (X) 40–48=–8 (Y) 42 – 46 = – 4
3	Squared up the deviations (d_x^2 and d_y^2)	(–8) (–8) = 64 (–4) (–4) = 16
4	Find the product of deviation ($d_x . d_y$)	(–8) (–4) = 32
5	Find the aggregates of d_x, d_y, d_x^2, d_y^2 and $d_x . d_y$	Σd_x = – 30 Σd_y = – 7 Σd_x^2 = 1860 Σd_y^2 = 2077 $\Sigma d_x d_y$ = 1862

6 Compute coefficient of correlation by the formula as explained below.

$$r = \frac{\frac{1}{N}\sum d_x d_y - \left(\frac{\sum d_x}{N}\right)\left(\frac{\sum d_y}{N}\right)}{\sqrt{\frac{\sum d_x^2}{N} - \left(\frac{\sum d_x}{N}\right)^2} \times \sqrt{\frac{\sum d_y^2}{N} - \left(\frac{\sum d_y}{N}\right)^2}}$$

$$r = \frac{\frac{1}{10} \times 1862 - \left(\frac{-30}{10}\right) \times \left(\frac{-7}{10}\right)}{\sqrt{\frac{1860}{10} - \left(\frac{-30}{10}\right)^2} \times \sqrt{\frac{2077}{10} - \left(\frac{-7}{10}\right)^2}}$$

$$r = \frac{186.2 - (-3) \times (-0.7)}{\sqrt{186 - (-3)^2} \times \sqrt{207.7 - (-0.7)^2}}$$

$$r = \frac{186.2 - 2.1}{\sqrt{177} \times \sqrt{207.21}}$$

$$r = \frac{184.1}{13.3 \times 14.39}$$

= 0.962 (Approx.)

Check Your Progress I

Compute product moment corelation for the following data:

X: 45, 55, 56, 58, 60, 65, 68, 70, 75, 80, 85

Y: 56, 50, 48, 60, 62, 64, 65, 70, 74, 82, 90

.....

.....

.....

.....

.....

Calculation of the Coefficient of Correlation by Rank Differences (Spearman's rho 'ρ')

One applies this method to calculate the coefficient of correlation when the direct quantitative measurement of the phenomenon under study is not possible, for example efficiency, intelligence etc. Pearson's coefficient of correlation between the ranks of two such traits is called the rank correlation between these traits. The formula for computing rank correlation is:

$$\rho = 1 - \frac{6\sum D^2}{N\left(N^2 - 1\right)}$$

Where ρ denotes coefficient of rank correlation between paired ranks, D denotes the difference between the paired ranks and N stands for the number of pairs. The following example would illustrate the above formula:

Table 3.4 : Ranking of Social Workers According to their Efficiency by Two Different Judges

Social Worker	Ranking by Judge A	Ranking by Judge B
A	3	4
B	9	7
C	6	6
D	5	8
E	1	1
F	2	3
G	4	2
H	7	5
I	8	10
J	10	9

Table 3.5

Social Worker	Ranking by Judge A (R_1)	Ranking by Judge B (R_2)	D = (R_1-R_2)	D^2
A	3	4	-1	1
B	9	7	+2	4
C	6	6	0	0
D	5	8	-3	9
E	1	1	0	0
F	2	3	-1	1
G	4	2	+2	4
H	7	5	+2	4
I	8	10	-2	4
J	10	9	+1	1
N = 10				$\Sigma D^2 = 28$

$$\rho = 1 - \frac{6\sum D^2}{N\left(N^2 - 1\right)}$$

$$\rho = 1 - \frac{6 \times 28}{10\left(10^2 - 1\right)}$$

$$\rho = 1 - \frac{168}{10\left(100 - 1\right)}$$

$$\rho = 1 - \frac{168}{10 \times 99}$$

$$\rho = 1 - \frac{168}{990} = 1 - 0.1696969 = 0.83 \text{ (approximately)}\chi$$

Chi-square Test

The χ^2 (Greek letter χ^2 and pronounced as Ki-square) test provides us with a method to evaluate whether or not frequencies, which have been empirically observed, differ significantly from those, which would be expected under a certain set of theoretical assumptions. For example, suppose categories of employees and preference of performance appraisal have been cross classified and the data summarized in the following 2 × 3 contingency table:

Table 3.6 : Relationship between Categories of Employees and Preference of Performance Appraisal

Nativity of Students	Preference of Performance Appraisal			Total
	Self	Immediate Supervisor	HOD	
Executives	38	20	12	70
Non-Executives	10	26	6	42
Total	48	46	18	112

The proportions of executives are 38/48 = 0.79, 20/46 = 0.34, and 12/18 = 0.67 (rounded to two decimals) for three raters. We would then want to know whether or not these differences are statistically significant.

To this end, we can assume null hypothesis: that there are no differences between categories of employees and preference of performance appraisal. This means that the proportions of executives and non-executives should be the same in each of the three types of appraisal. On the basis of the assumption that the null hypothesis is correct we can compute a set of frequencies that would be expected given these marginal totals. In other words, we can compute the number of executives who prefers self-appraisal and would be expected to be executives and compare this figure with that actually obtained. If the null hypothesis is true, we can compute a common proportion as:

$$\frac{38+20+12}{48+46+18}=\frac{70}{112}=0.625$$

With this estimate we would expect 48 × (0.625)=30 executives opting for self appraisal 46 × (0.625) = 28.75 executives opting for Immediate Supervisor and 18 × (0.625) = 11.25 executives opting for Head of the Dept. from 70 executives. Subtracting these figures from the respective size of the three samples, we find that 48 – 30 = 18 opting for self appraisal, 46 – 28.75 = 17.25 opting for Immediate Supervisor and 18 – 11.25 = 6.25 opting for Head of the Dept from 42 non-executives.

These results are shown in the following table, where expected frequencies are shown in parentheses.

Table 3.7

Categories of Employees	Preference of Performance Appraisal			Total
	Self	Immediate Supervisor	HOD	
Executives	38 (30)	20 (28.75)	12 (11.25)	70
Non-Executives	10 (18)	26 (17.25)	6 (6.75)	42
Total	48	46	18	112

To test the null hypothesis we compare the expected and observed frequencies. The comparison is based on the following χ^2 statistics.

$$\chi^2=\Sigma\left(\frac{(O-E)^2}{E}\right)$$

Where O stands for observed frequencies and E stands for expected frequencies.

Table 3.8 : Computation of χ^2

O	E	O – E	$(O - E)^2$	$\frac{(O-E)^2}{E}$
38	30.00	8.00	64.0000	2.1333
10	18.00	8.00	64.0000	3.5555
20	28.75	8.75	76.5625	2.6630
26	17.25	8.75	76.5625	4.4384
12	11.25	0.75	0.5625	0.0500
6	6.75	0.75	0.5625	0.0833
112	112.00			χ^2 = 12.9237

Before going into further details of the test the reader is introduced to certain terms used in this connection.

Level of Significance

As stated earlier, the Chi-square test is used to evaluate whether the difference between observed and expected frequencies is due to the sampling fluctuations and as such insignificant or whether the difference is due to some other reason and as such significant.

Before drawing the inference, that the difference is significant, researchers set up a hypothesis, often referred as a null hypothesis (symbolized as (H_0) as contrasted with the research hypothesis (H_1) that is set up as an alternative to H_0. Usually, although not always, the null hypothesis states that there is no difference between several groups or no relationship between variables then the probability of the occurrence of such a difference is determined. The probability indicates the extent of reliance that we can place on the inferences drawn. The table values of chi-square are available at various probability levels. These levels are called levels of significance. We can find out from the table the values of chi-square at certain levels of significance. Usually, the value of chi-square at 0.05 or .01 levels of significance from the given degrees of freedom is seen from the table and is compared with observed value of Chi-square. If the observed value of χ^2 is more than the table value at 0.05, it means that the difference is significant.

Degree of Freedom

To use the chi-square test, the next step is to calculate the degrees of freedom. Suppose we have a 2 × 2 contingency table like the one in Figure 3.1 given below.

	Column 1	Column 2	
Row 1	P	X	R_{t1}
Row 2	X	X	R_{t2}
	C_{t1}	C_{t2}	

Row Totals

Column Totals

Fig. 3.1

When one knows that the row and the column totals are $r_{t1,}$ r_{t2} and c_{t1} and $c_{t2.}$ respectively, then the number of degrees of freedom can be defined as the number of cell values that one can freely specify. In Fig. 3.1 given below, once you specify the one value of Row 1 (denoted by check in the figure) the second value in that row and the values of second row (denoted by X) are already determined; you are not free to specify these because you know the row totals and column totals. This shows that in a 2 × 2 contingency table you are free to specify one value only. The following example would illustrate the concept:

Table 3.9

	Affected A	**Non affected**	**Total α**
Inoculated	AB		
B	14 (20)	αB 26	40
Non inoculated	Aβ	αβ	
β	16	4	20
Total	30	30	60

Let us suppose that the two attributes A and B are independent then the expected frequency of the cell AB would be

Consequently, the frequencies of remaining three cells are automatically fixed. Thus for cell ab expected frequency must be 40 – 20 = 20, similarly for the cell aβ it must be 30 – 20 = 10 and for αβ it must be 10. This means that for 2 × 2 tables you have one choice of your own and in remaining three cells you have no freedom. Thus, degrees of freedom (*df*) can be calculated by the formula: ***df = (c – 1) (r – 1)***

Where *df* stands for the degrees of freedom, *c* for the number of columns and *r* for the number of rows. Thus in 2 × 3 table (above Table 3.9):

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

χ^2 (table value) (Level of significance 0.05, and degree of freedom 2) = 5.99

Inference

Since calculated value of χ^2 (=12.923669) is more than the table value of χ^2 (5.99) at 0.05 level of significance for 2 degree of freedom and hence the null hypothesis (H_0 = No difference among executives and non-executives with respect to their preference for performance appraisal) is rejected. We therefore, conclude that there is association between categories of employees (executives and non-executives) and their preferences for performance appraisal.

Regression Analysis

In this section, you will learn about how to calculate the regression line, using an equation that relates the two variables mathematically. The equation for a straight line where the dependent variable Y is determined by the independent variable X is:

Table 3.5

Social Worker	Ranking by Judge A (R_1)	Ranking by Judge B (R_2)	D = (R_1-R_2)	D^2
A	3	4	-1	1
B	9	7	+2	4
C	6	6	0	0
D	5	8	-3	9
E	1	1	0	0
F	2	3	-1	1
G	4	2	+2	4
H	7	5	+2	4
I	8	10	-2	4
J	10	9	+1	1
N = 10				$\sum D^2 = 28$

$$\rho = 1 - \frac{6\sum D^2}{N\left(N^2 - 1\right)}$$

$$\rho = 1 - \frac{6 \times 28}{10\left(10^2 - 1\right)}$$

$$\rho = 1 - \frac{168}{10\left(100 - 1\right)}$$

$$\rho = 1 - \frac{168}{10 \times 99}$$

$$\rho = 1 - \frac{168}{990} = 1 - 0.1696969 = 0.83 \text{ (approximately)}\chi$$

Chi-square Test

The χ^2 (Greek letter χ^2 and pronounced as Ki-square) test provides us with a method to evaluate whether or not frequencies, which have been empirically observed, differ significantly from those, which would be expected under a certain set of theoretical assumptions. For example, suppose categories of employees and preference of performance appraisal have been cross classified and the data summarized in the following 2 × 3 contingency table:

Table 3.6 : Relationship between Categories of Employees and Preference of Performance Appraisal

Nativity of Students	Preference of Performance Appraisal			Total
	Self	Immediate Supervisor	HOD	
Executives	38	20	12	70
Non-Executives	10	26	6	42
Total	48	46	18	112

So, we can substitute these values a and b into equation

Y = a + bX = 20 + 2 X

Using the estimation equation, we can predict what the academic performance score would be from the amount spent monthly on education. If the family spends 8 thousand, it can expect to score approximately 36 in academic performance:

Y = 20 + 2 (8)=20+16= 36

3.3 MEASURES OF DIFFERENCE

This section deals with parametric tests specially the 't' and 'f' tests. Parametric tests are based on the assumption that the population is normally distributed. In addition these tests use higher measurement levels, i.e. intervals and ratio.

t- Tests

There are two types of t –tests; one is called t –test for independent samples, the other one is called paired t-test. The first test is used for the scores of one group and is independent of the scores of the other group. That means, there are no logical relationships between the scores that have been obtained for one group when compared with other group. However, both the tests are used to assess significance of difference.

The Paired t – test

The data for this example is taken from an NGO in the social sector. The President wants to test the effects of a training programme for social workers to improve their performance. A group of 10 social workers are randomly selected from the NGO. A rating scale was administered before the training programme to assess the performance of social workers. The scale has a scoring range from 5 to 15. After the training programmes the same rating scales was administered. The pre-training and post- training scores are shown in the Table 3.12.

The President establishes a null hypothesis that there is no difference between pre-training and post- training scores. The research hypothesis is that post- training scores will show improvement in performance over pre- training.

Table 3.12 : Performance Scores of 10 Social Workers

Social Workers (X)	Pre-Training Scores (Y)	Post-Training Scores	D=(Y–X)	D^2
1	9	12	3	9
2	8	10	2	4
3	15	15	0	0
4	12	14	2	4
5	8	14	6	36
6	4	11	7	49
7	6	10	4	16
8	3	8	5	25
9	3	8	5	25
10	2	8	6	36
N = 10			$\Sigma D = 40$	$\Sigma D^2 = 204$

The procedure for calculating 't' is given below:

Table 3.13

Step	Procedure	Application to above Table
1	Find out the difference (D) between the post training and pre training scores	D = Y – X = 12 – 9 = 3 10 – 8 = 2 ……
2	Compute Mean Difference	$\bar{X} = \frac{\sum D}{N} = \frac{40}{10} = 4$
3	Compute square of difference	$D^2 = (3)^2 = 9, (2)^2 = 4$ . . . . .
4	Find the sum of squares of difference	$\sum D^2 = 204$
5	Compute sum of squares (SS)	$SS = \sum D^2 - \frac{(\sum D)^2}{N}$ $= 204 - \frac{(40)^2}{10} = 204 - 160 = 44$
6	Find out degree of freedom(df)	df = n – 1 = 10 – 1 = 9
7	Compute variance (S^2)	$S^2 = \frac{SS}{df} = \frac{44}{9} = 4.89$
8	Compute 't' by using the formula given above	$t = \frac{\bar{X} - \mu}{\sqrt{S^2/N}} = \frac{4-0}{\sqrt{4.89/10}}$ $= \frac{4}{0.489} = 5.7225$

Now that the value of the paired t – test has been calculated, we have to see if the null hypothesis can be rejected. It is assumed that the, training is likely to improve the post- training scores. That means there is directionality in the data. Hence we will use one-tailed test. As such, we refer to critical value (Refer a text book on Statistics). Looking at the .05 level of significance for one tailed test we go down to the intersect of 9 degree of freedom (df). The calculated value of 't' (5.7225) is greater than the critical value (1.833), hence, the null hypothesis can be rejected and we can say that there is statistically significant difference in social workers performance after the training.

The t–test for Independent Samples

The t – test for two independent samples examines the difference between their means to see how close or apart they are. Let us consider the data given in the Table 3.13. The data compares performance of social workers working on two programmes.

President of the organisation is interested in studying whether there is significant difference among the performance of social workers. The data at the interval level of measurement are presented in the Table 3.13.

Let us call the performance of social workers of programme A as X and the performance of social workers of programme B as Y. The first step in calculating t – test for two independent samples is to square the values of X and Y columns to get X^2 and Y^2 values.

The next step is to sum up the columns to find out $\sum X$, $\sum Y$, $\sum X^2$ and $\sum Y^2$. This is followed by calculation of means of X and Y columns, that is X and Y.

Table 3.14

S.No.	Performance of social workers of programme A (X)	X^2	Performance of social workers of programme B (Y)	Y^2
1	8	64	12	144
2	11	121	9	81
3	9	81	6	36
4	12	144	5	25
5	16	256	8	64
6	10	100	12	144
7	7	49	11	121
8	16	256	10	100
9	6	36	10	100
10	5	25	7	49
N=10	**ΣX= 100**	**ΣX^2 = 1132**	**ΣY= 90**	**ΣY^2=864**

The procedure for calculating 't' is given below;

Table 3.15

Step	Procedure	Application to Previous Table
1.	Square the scores of X and Y column to get X^2 and Y^2 values.	= $(8)^2$ = 64, $(11)^2$ = 121, $(9)^2$ = 81,.. = $(12)^2$ = 144, $(9)^2$ = 81, $(6)^2$ = 36,..
2.	Find the sums of columns X, Y, X^2 and Y^2	ΣX = 100 ΣY = 90 ΣX^2 = 1132 ΣY^2 = 864
3.	Compute Mean Scores for column X and Y	$\overline{X} = \frac{100}{10} = 10$
4.	Compute Variance sum of square for X and Y column (SS_x and SS_y)	$= 1132 - \frac{(100^2)}{10}$ = 132
5.	Find out the sum of squares	$SS_c = SS_x + SS_y$ = 132 + 54 = 186
6.	Find out the degrees of freedom for two sets of data (df_c)	$df_x = n - 1 = 10 - 1 = 9$ $df_y = n - 1 = 10 - 1 = 9$ $df_c = df_x + df_y = 9 + 9 = 18$

7.	Compute combined variances S^2_c	$S^2_c = SS_c/df_c = 186/18 = 10.33$
8.	Compute 't' by using the formula	$t = \frac{\bar{X} - \bar{Y}}{S_c\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$ $t = \frac{10-9}{\sqrt{10.33} \times \sqrt{\frac{1}{10} + \frac{1}{10}}}$ $t = \frac{1}{3.214 \times 0.4472} = \frac{1}{1.43} = 0.699$

Before we take the decision we have to check whether the performance scores of social workers on two programmes show any directionality. Since there is no indication that either set of data has influence over other, hence this is non-directional hypothesis. Thus, we will have to look for critical value of t for two tailed tests at 0.05 level of significance for 18 degree of freedom. (Refer a text book on Statistics) to find the critical value of 't', which is equal to 2.101. Since, our calculated value 0.699 is not larger than the critical value of 2.101, the null hypothesis cannot be rejected. That means there is no significant difference between the performance of SocialWorkers in two programmes.

The One-Way Analysis of Variance

The analysis of variance (some times obtained as ANOVA) determines whether there is a statistically significant difference between more than two sets of data on the basis of their means. There are two variances that are calculated in F – test procedure. The first variance is obtained from between the groups and the second variance is obtained from the values within each of the groups. Let us take the number of cases handled by three social workers of an organisation. Chief functionary of the organisation wants to see whether there are significant differences between the performances of three social workers.

Table 3.16 : Number of Cases Handled by Three Social Workers

Month	No. of Cases handled by Social Worker A	A^2	No. of Cases handled by Social Worker B	B^2	No. of Cases handled by Social Worker C	C^2
January	14	196	38	1444	34	1156
February	20	400	42	1764	40	1600
March	22	484	24	576	30	900
April	18	324	12	144	48	2304
May	30	900	40	1600	58	3364
	ΣA = 104	**ΣA^2= 2304**	**ΣB = 156**	**ΣB^2= 5528**	**ΣC = 210**	**ΣC^2= 9324**

$n_1 = 5,\ n_2 = 5,\ n_3 = 5,\ n = n_1 + n_2 + n_3 = 5 + 5 + 5 = 15$

$\Sigma T = \Sigma A + \Sigma B + C = 104 + 156 + 210 = 470$

A null hypothesis is formulated stating that there are no differences between the performances of three social workers. The procedure of calculating 'f' test is as under;

Table 3.17

Step	Procedure	Application to Table 13
1.	Find out the squares of scores in each of the three columns	$A^2 = (14)^2 = 196,\ (20)^2 = 400,\ \ldots.$ $B^2 = (38)^2 = 1444,\ (42)^2 = 1764,\ldots$ $C^2 = (34)^2 = 1156,\ (40)^2 = 1600,\ldots..$
2.	Find the sums of values in each of the three columns & similarly sum of squares	$\Sigma A = 104 \quad \Sigma A^2 = 2304$ $\Sigma B = 156 \quad \Sigma B^2 = 5528$ $\Sigma C = 210 \quad \Sigma C^2 = 44100$
3.	Add the sums of the values of each of the three columns	$\Sigma T = \Sigma A + \Sigma B + \Sigma C$ $(\Sigma T) = 104 + 156 + 210 = 470$
4.	Compute ΣT^2	$\Sigma T^2 = \Sigma A^2 + \Sigma B^2 + \Sigma C^2$ $= 2340 + 5528 + 44100 = 51968$
5.	Compute sum of square of total scores (SST)	$SST = \sum T^2 - \left(\frac{(\sum T)^2}{n} \right)$ $= 51968 - 14726.667 = 37241.333$
6.	Compute sum of squares between the scores (SSB)	$SSB =$ $SSB =$ $\frac{(104)^2}{5} + \frac{(156)^2}{5} + \frac{(210)^2}{5} - \frac{(470)^2}{15}$ $= 2163.2 + 4867.2 + 8820 - 14726.667$ $= 15850.4 - 14726.667 = 1123.733$
7.	Compute Error SS_W = SST – SSB	$= 37241.33 - 1123.733 = 36117.597$
8.	Prepare a summary table	Table 15

Table 3.18 : Summary Table

Source of Variation	df	SS	Mean Squares
Between measurement	K–1=3–1=2	1123.733	561.8665
Within measurement	N–K=15–3=12	36117.597	3009.8
Total	N–1=15–1=14	37241.33	

K=Number of Groups

N=Total Number of Observations

$$F\ Ratio = \frac{SS_B}{SS_W}$$

$$F_{2,12} = \frac{1123.733}{36117.597} = 0.031$$

Since calculated value of F is not larger than the critical value (Refer a text book on Statistics) we cannot reject the null hypothesis. Hence, we infer that there is no statistically significant difference between the performances of three social workers.

3.4 TESTING OF HYPOTHESIS

The Steps in Testing Hypothesis

1) State the Research Hypothesis (H_1)
2) Formulate the Null Hypothesis (Ho)
3) Choose a statistical Test
4) Specify a significance level.
5) Compute the statistical test
6) Reject/accept the Ho
7) Draw the inference i.e. accept/reject H_1

Step 1 : State the Research Hypothesis (H_1)

H_1 Employees' salary and level of job satisfaction are likely to be interdependent.

Step 2 : Formulate the Null Hypothesis (Ho)

Ho Employees' salary and level of job satisfaction are independent.

Step 3 : Choose a statistical Test

Let us suppose that we have decided to use Chi-Square statistic (χ^2) to test the interdependence between the variables considered in the research hypothesis.

Step 4 : Specify a significance level

Further we suppose that we would like to test our hypothesis at .05 level of significance.

Step 5 : Compute the statistical test.

In this step the researcher has to cross-tabulate his data and compute chi-square test. On computation of the test, let us say that the test yielded a value of 6.78, df = 1.

Step 6 : Reject/accept the Ho

Since the calculated value of chi-square is more than the critical value we reject the null hypothesis .

Step 7 : Draw the inference i.e. accept/reject H_1

We accept the research hypothesis because the null hypothesis has been rejected. Hence, we can infer that employees' salary and level of job satisfaction are interdependent.

3.5 LET US SUM UP

Product Moment correlation and rank-difference correlation are the commonly used measures of relationship between any two variables. When the data are available in ordinal (rank) form of measurement and the size of the sample is small, then we compute rank-difference correlation. You have studied bivariate distributions and learnt methods to find relationship between two characterstics. Finally, you have endeavoured into inferential methods and learnt about chi-square test, t-tests and F-test. This chapter will thus provide you with a base in your pursuit in Statisitcs.

3.6 FURTHER READINGS AND REFERENCES

Alston, Margaret. and Bowles, Wendy. Research for Social Workers - an introduction to methods. Rawat Publications. New Delhi & Jaipur (2003).

Baker, L. Therese .*Doing Social Research,* New York: McGraw Hill, 1988.

Black, Allen & Babbie E. *Research Methodology for Social Work*, Belmont, California, Wadsworth, 1989.

Blalock, H.M.*Social Statistics*, Auckland. Mc Graw Hill, 1988.

Elhance, D.N. *Fundamental of Statistics*, Allahabad, Kitab Mahal, 1984.

Festinger. L. and Katz. D. (eds.) *Research Methods in the Behavioural Sciences,* New York : The Dryden Press,1953.

Freud, J.E. *Modern Elementary Statistics*, New Delhi, Prentice Hall, 1977.

Garret, H.E. *Statistic in Physical & Education*, Bomb: Vakils, 1966.

Gupta, S.P. *Statistical Methods*, New Delhi, S.Chand, 1980.

Kidder, Lauise H. *Research Methods in Social Relations,* New York: Holt, 1981.

Krishef, Curtis H. Fundamentals Statistics for Human Services and Social Work, Boston: Duxbury Press, 1987.

Lal Das, D.K. (2000), Practice of Social Research: A Social Work Perspective, Rawat Publications, Jaipur.

Levin, Richard I. *Statistics for Management* Delhi : Pearson Prentice Hall, 2005.

Monette, Duana R. et. al. *Applied Social Research: Tool For the Human Services,* Chicago: Holt, 1986.

Nachmias D and Nachmias C. *Research Methods in the Social Sciences*, New York: St. Martins press, 1981.

Ostle, B. and Mensing R.W. *Statistics in Research*, New Delhi, Oxford, 1975.

Philip, A.E. et. al. *Social Work Research*, and The Analysis of Social Data, Oxford: Peragon Press, 1975.

Sanders, D.H. et. al. *Statistics A Fresh Approach*, New Delhi: Mc Graw Hill, 1975.

Thyer, Bruce. The Handbook of Social Work Research Methods. Sage Publications Inc. (USA). (2010)

UNIT 4 REPORTING OF RESEARCH

Structure

4.0 OBJECTIVES

After the completion of this Unit, you should be able to:

- state the reasons for writing a research report;
- list the three main components of a research report;
- describe each component of a research report; and
- write the final report of any research study, viz., its beginning, the main body, and the end of the report.

4.1 INTRODUCTION

Every research activity is concluded by presenting the results and discussions. The reporting of a research study depends on the purpose with which it was undertaken. One might have conducted a study as a personal research, as an institutional project, as a project funded by an outside agency or towards fulfilling the requirement for the award of a degree.

Research studies, when reported, follow certain standard patterns, styles and formats for maintaining parity in reporting and for easy grasp by others who are concerned with those studies. The present chapter is devoted to this aspect of research: How to write a research report? It starts with the objectives of writing a research report followed by the components of the report itself (the beginning, the main body and the end).

4.2 WHY AND HOW TO WRITE A RESEARCH REPORT

Once you complete your research project, you are expected to write the report. A research report is a precise presentation of the work done by a researcher while investigating a particular problem. Whether the study is conducted by an individual researcher, by a team or by an institution, the findings of the study should be reported for several reasons. These are:

- People learn more about the area of study,
- The discipline gets enriched with new knowledge and theories,
- Researchers and practitioners in the field can apply, test and retest the findings already arrived at,
- Other researchers can refer to the findings and utilise the findings for further research and
- Findings can be utilized and implemented by the policy makers or those who had sponsored the project.

The final report of a research exercise takes a variety of forms.

- A research report funded by an educational institution may be in the form of a written document.
- A research report may also take the form of an article in a professional journal.
- The research reports of students of M.A., M.S.W., M.Sc., M.Ed., M Phil. or Doctoral Programme take the form of a thesis or dissertation.

In the following sections we shall discuss the main components of a research report. The entire research report is mainly divided into three major divisions: the beginning, the main body and the end (please see box).

Beginning	Main Body	End
• Cover/Title Page	• Introduction	• Bibliography/ References
• Second Cover	• Review of Literature	
• Acknowledgements	• Research Methodology	• Appendices
• Contents	• Analysis and Interpretatiopn of Data	
• List of Contents	• Main Findings and Conclusions	
• List of Tables	• Summary	

4.3 THE BEGINNING

The beginning of a report is crucial to the entire work. The beginning or the preliminary section of the research report contains the following items, more or less in the order given below:

- Cover or Title Page
- Acknowledgements
- Table of Contents
- List of Tables
- List of Figures and Illustrations
- Glossary

Let us describe in brief each of the above six items of the preliminary section of a report.

i) Cover or Title page

The cover page is the beginning of the report. Though different colleges, universities and sponsoring institutions prescribe their own format for the title page of their project report or thesis, generally, it follows the downward vertical order:

- title of the topic,
- relationship of the report to a degree, course, or organisational requirement,
- name of the researcher,
- name of the supervisor,
- name of the institution where the report is to be submitted, and
- year or month of submission.

The title page should carry a concise and adequately descriptive title of the research study. The title should briefly convey what the study is about. Researchers tend to make errors in giving the title by using too many redundant and unimportant words.

Here, we have drawn a list of a few titles of research reports and doctoral theses:

a) A Critical analysis of Textual Material for Principles of Accounting and its Translation for Distance Education

b) Developing Self-Instructional Material

c) Planning Design and Development of one Self-Instructional Unit in print

In title (b), it is not clear at which level the researcher is developing self-instructional material.

The title should be written either in bold letters or upper-lower case and should be placed in the central portion of the top of the cover page. Here, we have reproduced the cover page of a research report in Box 1.

Box 1 : Example of the Title Page of a Research Report

Evaluation Scheme of Assistance to Organisations for the Disabled
Sponsored By:
Ministry of Social Welfare Government of India
Director
Dr. D.K. Lal Das
Principal
College of Social Work (ICSW)
Red Hills, Hyderabad - 500004
March, 2002

Note the other points mentioned on the cover page. Also, observe the placement of these points.

ii) Preface or/and Acknowledgement

Preface is not a synonym for either a Acknowledgement or a Foreword. A preface should include the reasons why the topic was selected by the researcher. It may explain the history, scope, methodology and the researcher's opinion about the study. The preface and acknowledgements can be in continuation or written separately. This page follows the inner title page. It records acknowledgement with sincerity for the crucial help received from others to conduct the study. The acknowledgement should be non-emotional and simple.

iii) Table of Contents

A table of contents indicates the logical division of the report into various sections and subsections. In other words, the table of contents presents in itemized form, the beginning, the main body and the end of the report. It should also indicate the page reference for each chapter or section and sub-section on the right hand side of the table.

Sample table of contents is given below:

Box 2 : Table of Contents

iv) **List 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 the page number on which the table has appeared. We provide you in Box 3 an example of a list of tables.

Box 3: Example of a List of Tables

Table	Title	Page
1	Population of Disabled	2
2	Age of the Beneficiaries	18
3	Level of Education	20
4	Sex of the Beneficiaries	21
5	Occupation of the Beneficiaries	22
6	Caste of the Beneficiaries	23
7	Marital Status	24
8	Size of the Family	25
9	Income of the Family	26
10	Nature of the Disability	27
11	Age of Non-Beneficiaries	28
12	Level of Education of Resident Beneficiaries	30
13	Sex	31
14	Caste	32
15	Marital Status	33
16	Size of the Family	34
17	Income of the Family	35
18	Nature of Disability	37
19	Staffing Pattern	46
20	Year-wise Distribution of grants-in-aid	51
21	Age and Opinion of Beneficiaries on Utility of the Services Provided under the Scheme	59

v) **List of Figures and Illustrations**

The page' List of Figures' comes immediately after the 'List of Tables' page. You will observe in the following example that the list of figures is written in the same way as the list of tables.

Box 4: Example of List of Figures

vi) **Glossary**

A glossary is a short dictionary, explaining the technical terms and phrases which are used with special connotation by the author. Entries of the technical termed are made in alphabetical order. A glossary may appear in the introductory pages although it usually comes after the bibliography. An exemplar glossary is given below.

Algorithm : A step-by-step procedure consisting of mathematical and/or logical operations for solving a problem.

Artificial Intelligence (AI): The study of computer techniques that mimic certain functions typically associated with human intelligence.

Back-up : Duplication of a program or file on to a separate storage medium so that a copy will be preserved against possible loss or damage to the original.

Benchmark : A measured point of reference from which comparisons of any kind may be made, often used in evaluating hardware and software, in comparing them against one another.

Command : An instruction to the computer which is not a part of a program.

Cybernetics : The field of science involved in comparative study of the automatic control or regulation of and communication between machine and man. These studies include comparisons between information-handling machines and the brains and nervous systems of animals and humans.

Data : Input in to a computer that is processed by mathematical and logical operations so that it can ultimately become an intelligible output.

Data Processing : The input, storage, manipulation and dissemination of information using sequences of mathematical and logical operation.

Electronic Spreadsheet : Software that simulates a worksheet in which the user can indicate data relationships. When data are changed, the program has the ability to instantly recalculate any related factors and save all the information in memory.

Graphics Package : A programme that helps draw graphs.

Hard Copy : Output of information in permanent form, usually on paper, as opposed to temporary display on a CRT screen.

Ink Jet Printer : A type of printer in which dot matrix characters are formed by ink droplets electrostatically aimed at the paper surface.

Laser Printer : A printer that uses a laser beam to form images on photo-sensitive drums. Laser printers are now used as output devices for computers.

Megabyte (MB or M-Byte) : A unit of data storage in a computer, such as 1024 kilobytes or 10241024 bytes.

Personal Computer : A moderately priced computer, designed principally for a single user in a home or small-office environment.

Source: Balagurusamy E: Selecting and Managing Small Computers.

ii) **List of Abbreviations Italics**

To avoid repeating long names again and again, a researcher uses abbreviations. Since abbreviations are not universal, it is necessary to provide the full form of the abbreviations in the beginning. An exemplar list of abbreviation is given below.

Abbreviations

AIMA	All India Management Association
AIR	All India Radio
APPEP	Andhra Pradesh Primary Education Project
AVRC	Audio Visual Resource Centre
ATI	Administrative Training Institute
BEL	Bharat Electronics Limited
BEO	Block Education Officer
BRC	Block Resource Center
BSE	Board of Secondary/Senior Secondary Education

CABE	Central Advisory Board of Education
CBT	Computer-Based Training
CEO	Circle Education Officer
CIET	Central Institute of Educational Technology
CRC	Cluster Resource Center
CSS	Centrally Sponsored Scheme
DIET	District Institute of Educational Technology
DIT	District Institute of Training
DD	Doordarshan
DOE	Department of Electronics
DoSpace	Department of Space
DOT	Department of Telecommunication
DPEP	District Primary Education Program
DPEPII	District Primary Education Program: Phase II
EMRC	Education Media Resource Centre

Activity

Take any report which has been prepard by your institution and check whether the title page contains all the essential information. If not, try to fill in the gaps.

.....

.....

.....

.....

.....

Check Your Progress I

Note: Use the space given below for your answers.

1) List the major parts of 'the beginning' of a research report. Describe briefly the importance of each part.

.....

.....

.....

.....

.....

.....

4.4 THE MAIN BODY

The main body of the report presents the actual work done by an investigator or a researcher. It tells us precisely and clearly about the investigation/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. The report categorically avoids unnecessary details and loose expressions; we shall examine this point in detail in this section. The table of contents for the report outlined of six section/chapters in the main body. These are:

- Introduction
- Review of Literature
- Research Methodology
- Analysis and Interpretation of Data
- Major Findings
- Conclusions and Discussions
- Summary

Besides the logicality of sections/chapters in the main body, there are certain other important aspects which need close attention. These are the style of writing, the design and placement of references and footnotes, the typing of the report and the tables and figures.

Let us elaborate these points in the following sub-sections.

Chapters and their Functions

We will discuss the chapterisation of a thesis or a research report under six heads as noted above. Let us begin with introduction which is usually the first chapter.

Introduction

This is the first chapter of a research report. It introduces the topic or problem under investigation and its importance. The introductory chapter:

- Gives the theoretical background to the specific area of investigation,
- States the problem under investigation with specific reference to its placement in the broader area under study,
- Describes the significance of the present problem,
- Defines the important terms used in the investigation and its reporting,
- States precisely the objective(s) of the study,
- States precisely the important terms used in the study that would be tested through statistical analysis of data, and
- Defines the scope and limitations of the investigations.

Although these sub-sections are common, it is not necessary to follow the given order strictly; there may be variation in the order of the sub-sections. Sometimes the review of literature related to the area under investigation is also presented in

the first chapter and is placed immediately after providing the theoretical background to the problem. Many researchers use review to argue the case for their own investigation. In experimental research it becomes essential to review related studies to formulate the hypotheses.

Review of Literature

The second chapter of a research report usually consists of the review of important literature related to the problem under study. This includes the abstraction of earlier research studies and the theoretical articles and papers of important authorities in the field. This chapter has two functions. Firstly, while selecting a problem area or simply a topic for investigation, the researcher goes through many books, journals, research abstracts, encyclopaedia, etc. to finally formulate a problem for research in order to decide on a specific problem for investigation. The review of literature is the first task for a research in order to decide on a specific problem for investigation. It also helps in formulating the theoretical framework for the entire study. Secondly, such a review 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.

Besides these, the review also indicates the researcher's grasp over the area under investigation, and also his/her efficiency to carry out the study. While reviewing literature in the area concerned, you have to keep in mind that (reviewed) literature has to be critically analysed and summarized in terms of agreements and disagreements among the authors and researchers in order to justify the necessity for conducting your investigation. Researchers may make two types of errors in their review exercises. Many researchers simply report the findings of one study after another in a sequential order without showing how the findings are connected with one another. Others report on studies that are at best only marginally related to their own hypothesis.

The design of a study is usually described in the third chapter of the report. Broadly speaking, this chapter provides a detailed overview of "how" the study was conducted. The various sub-section include:

i) description of the research methodology,

ii) variables: the dependent, independent and intervening variables with their operational definitions;

iii) Sample: defining the population, and the sampling procedure followed to select the sample for the present study;

iv) listing and describing various tools and techniques used in the study, like questionnaires, attitude scales, etc., whether these have been adopted or developed by the investigator, their reliability, validity, item-description, administration and scoring, etc.;

v) describing the statistical techniques used in the analysis of data including the rationale for the use and method of data analysis.

Analysis and interpretation of data

This is fourth chapter of the research report. It is the heart of the whole report, includes the outcome of the research. The collected data are presented in as tabular form and analysed with the help of statistical techniques–parametric and/

or non-parametric. The tables are interpreted and if necessary, the findings are also presented graphically. The figures do not necessarily, repeat the tables, but present data visually for easy understanding and easy comparisons. Data may be presented in parts under relevant sections. The analysis of the data not only includes the actual calculations but also the final results. It is essential that at each stage of analysis the objective(s) of the study and their coverage is taken care of. This chapter also presents the details about the testing of each hypothesis and the conclusions arrived at. This gives the reader a clear idea regarding the status of the analysis and coverage of objectives from point to point.

Main findings and conclusion

This is usually the fifth chapter in a research report. The major findings of the study analysed and interpreted in the preceding chapter are precisely and objectively stated in this chapter. The fourth chapter contains such presentations as only a specialist or a trained researcher can understand because of the complexities involved; but in the fifth chapter the major findings are presented in a non-technical language so that even a non-specialist such as a planner or an administrator in the field can make sense out of them.

The main findings are followed by a discussion on the results/findings. The major findings are matched against the findings of other related research works which have already been reviewed in the second chapter of the report. Accordingly, the hypotheses formulated in the first chapter are either confirmed or discarded. In case the null-hypotheses are rejected, alternative hypotheses are accepted. If the findings do have any discrepancy in comparison with those of other researches or if the findings do not explain sufficiently the situation or problem under study or if they are inadequate for generalisations, explanations with proper justification and explanation have to be provided.

The implications should suggest activities for and provide some direction to the practitioners in the field. Unless these implications are clearly and categorically noted, it becomes difficult for the practitioners to implement them on the one hand, and on the other, research findings do not get utilised at all even if they have been recorded in a report.

The implications follow a presentation/listing of the limitations of the study on the basis of which suggestions are made to carry out further investigation or extend the study from where it has reached.

Summary

Some researchers include a summary alongwith the research report (as the last chapter) or as a pull-out to the report itself. If sums up precisely the whole of the research report right from the theoretical background to the suggestions for further study. Sometimes researchers get tempted to report more than what the data say. It is advisable to check this tendency and be always careful to discuss the report only within the framework provided by the analysis and interpretation of data, i.e., within the limits of the findings of the study.

Check Your Progress II

Note: Use the space given below for your answers.

1) Comment briefly on the uses of (a) review of literature, and (b) conclusion in a research report.

.....

.....

.....

.....

.....

.....

Writing Style

The style of writing a research report is different from other writings. The report should be very concise, unambiguous, and presented meaningfully. The presentation should be simple, direct and in short sentences. Special care should be taken to see that it is not dull and demotivating.

Statements made should be as precise as possible – they should be objective and there should be no room for subjectivity, personal bias and persuasion. Similarly, over generalisation must by avoided. There is no place for hackneyed, slang and flippant phrases and folk expressions. The writing style should be such that the sentences describe and explain the data but do not try to convince or pursuade the reader. Since the report describes what has already been completed, the writing should be in the past tense.

In the 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., etc. should the avoided. Some authors recommend that the use of personal pronouns like "I", "We" etc., should be avoided, however, there is no hard and fast rule in this case. Similarly, a large number of research reports use passive voice which is strongly discouraged by the linguists. Similarly, abbreviations of words and phrases – like IGNOU, DDE, NIRD, etc. – should be used to avoid long names repeatedly inside the text, as well as in figures, tables and footnotes.

Special care should be taken while using quantitative terms in a report, such as *few* for number, *less* for quantity etc. No sentence should begin with numericals like "40 students", instead it should start as "Forty students". Commas should be used when numbers exceed three digits –1,556 or 523,489, etc.

Language, grammar and usage are very important in a research report, the *Roget's Thesaurus Handbook of Style* by Campbell and Ballon (1974), and a good dictionary would be of much help. MS-Word software provides good support to

— Spelling and Grammar

— Thesaurus

— Auto Correct

— Auto Summarise

A researcher is advised to use these features on the MS-Word to make the report error-free. It is always advisable 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 – even experienced researchers with many publications revise their reports many times before giving them for final typing.

References

Articles, papers, books, monographs, etc. quoted inside the text should always accompany relevant references, i.e., the author and the year of publication e.g., (Mukherjee, 1998). If a few lines or sentences are actually quoted from a source, the page number too should be noted e.g., (Mukherjee, 1998:120-124). Besides, full reference should be placed in the Reference section of the report (see section Bibliography and References).

In preparing the references, another factor to be considered is the abbreviations of words and expressions and their right placement. While writing a research report, abbreviations may be used to conserve space in references. If a researcher is not familiar with the abbreviations, he/she should consult the relevant literature as and when required. In the following table a comprehensive list of abbreviations has been given for ready reference (the Latin abbreviations have been italicised).

Table 4.1 : List of some important Abbreviations used in Footnotes and Bibliographies

Words	Abbreviation
About (approximate)	c. (*circa*)
Above	*supra*.
And the following	*et seg*.
And the following	*f.,ff*.
And others	*et. al*.
Article, articles	art., arts.
Article, articles	*infra*.
Book, books	bk., bks.
Chapter, chapter	chap., chaps.
Column, columns	Col., Cols.
Compare	*cf.*
Division, division	div., divs.
Editor, editors	ed., eds.
Edition, editions	ed., eds.
For example	*e.g.*
Figure, figures	fig., figs.
Here and there (scattered)	*passim*

Illustrated	Ill
Line, lines	l. ll.
Manuscript	ms.
Mimeographed	mimeo.
No date given	n.d.
No name given	n.n.
No place given	n.p.
Number, numbers	no., nos.
Page, pages	p., pp.
Part, parts	pt., pts.
Paragraph in length	(….)
Paragraph, Paragraphs	par., pars
Previously cited	*op. cit*.
Revised	rev.
Same person	*idem*.
Same reference	*ibid.*
Section, sections	sec., secs.
See	*vide*
The place cited	*loc. cit*
Thus	*sic*.
Translated	trans.

Typing of dissertations, research reports, project reports etc. needs greater care than other typed documents. In a research report, one does not expect overwriting, strikeovers, erasures and insertions.

Before typing the report, it is necessary to check whether the handwritten report, i.e., the manuscript is in a proper shape. Whether the manuscript of the report is typed by a typist or by the researcher himself/herself, a clear and comprehensible manuscript makes typing easy. Too many additions and corrections make the manuscript crammed. And a crammed manuscript makes typing difficult and time consuming. Only one side of the paper should be typed and typing should be double spaced. Space should be left on each side of the paper as follows:

- left side margin
- right side margin
- top margin
- bottom margin

If there is a lengthy quotation, it should be indented and typed in single space. At the end of each line, words should be divided as per convention. A dictionary which shows syllabification should be consulted if words are to be broken at all. Unlike the lengthy quotations, short quotation of three/four lines may be included in the text within quotation marks.

Subject to access to a computer and word processing software, it is better to prepare the report on a computer. It has several advantages, for example, you can

- edit time and again without incorporating new errors which is what happens when you use a manual typewriter,
- define your margin – top, bottom, left and right easily,
- define pages in landscape or portrait size, particularly for tables and diagrams,
- choose out of about 70+ fonts, shapes of letters and type-sizes from the smallest 8 point to the large 72 point,
- check spelling, grammar, synonyms and antonyms,
- choose illustrations from the clip-art file, and
- can index (alphabetical order) the references automatically.

Tables and Figures

Tables: Preparation and appropriate placement of tables in the text are equally important. They need careful attention from the researcher. Tables help the readers to get a quick view of the data and comprehend vast data at one go. However, tables should be presented only when they are necessary. Too many tables may confuse the reader, instead of facilitating his/her reading. As such you need to be selective in placing tables in the report. If data are too complicated to be presented in one table, several tables may be used to give a clear picture of the data in proper sequential order. Tables, if small, may accompany the textual material, and if large, should be put on one full page without mixing them with the text. All the tables should be numbered serially in the text, so that they may be quoted or referred to with the help of those numbers conveniently.

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, tables can be typed in smaller fonts like 8 point or 9 point to accommodate them on the same page. 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. The title of the table should be brief and self-explanatory.

Figures: Figures are necessary when the data is to be presented in the graphic form. They include charts, maps, photographs, drawings, graphs, diagrams, etc. The important function of a figure is to represent the data in a visual form for clear and easy understanding. Textual materials should not be repeated through figures unless very necessary.

Figures should be as simple as possible and the title of each figure should precisely explain the data that has been presented. Usually, a figure is accompanied by a table of numerical data. Again, figures are presented only after textual discussion and not the other way round. The title design of figures should be followed

consistently throughout the report. Every first letter of a word of the title should be in capitals, and figures should be numbered in Indian numerals like 1, 2, 3 etc. And the title, unlike for tables, is presented below the figure.

4.5 THE END

The end of the report consists of references and appendix/appendices. References come at the end after the last chapter of the report. The last section labelled as references appears at the top of a new sheet of paper. The reference section is a list of the works that have been cited in the report/thesis. All references quoted in the text are listed alphabetically according to the last name of the authors. The work of the same author should be listed according to the date of publication with the earliest appearing first.

Bibliography and References

Research reports present both bibliographies and references. Although many researchers use these terms interchangeably, the two terms have definite and distinct meanings. A bibliography is a list of titles – books, research reports, articles, etc. that may or may not have been referred to in the text of the research report. References include only such studies, books or papers as have been actually referred to in the text of the research report. Whereas research reports should present references, books meant for larger circulation may be listed in bibliographies that should include all such titles as have been referred to.

There are mainly two style manuals detailing general form and style for research reports. These are:

- American Psychological Association, *Publication Manual,* 3rd ed. Washington, DC: American Psychological Association, 1983.
- *The Chicago Manual of Style, 13th* rev.ed., Chicago University of Chicago Press, 1982.

Style of Referencing

There are mainly two types of referencing:

1) arranging references in alphabetical order where the researcher has cited the name of the author and year of publication/completion of the work in the text.

2) arranging references in a sequence as they appear in the text of the research report. In this case, related statement in the body of the text is numbered. However, most research reports use alphabetical listing of references.

For example, entries in a reference section may look like the following:

Gannicott, K. and Throsby, D. (1994), *Educational Quality and Effective Schooling*, UNESCO, (Book), Paris.

Koul, B.N., Singh, B. and Ansari, M.M. (1988), *Studies in Distance Education*, IGNOU & AIU, New Delhi.

Kumar, K.L. (1995), *Educational Technology*, New Age Publishers, New Delhi.

Ministry of Human Resource Development, DPEP: *Guidelines* (1995), Department of Education MHRD, Government of India, New Delhi.

Mukhopadhyay, M. (ed.) (1990), *Educational Technology: Challenging Issues*, Sterling Publishers, (Edited Book), New Delhi.

Mukhopadyay, M. (1998), "Teacher Education and Distance Education: The Artificial Controversy", in Buch, Piloo M., (ed.) *Contemporary Thoughts on Education*, SERD, (Chapter in Book), Baroda.

Parhar, M. (1993), Impact of Media on Student Learning, Unpublished Doctoral Dissertation, Jamia Millia Islamia, (Thesis), New Delhi.

Sachidananda, Tribal Education: New Perspectives and Challenges, *Journal of Indian Education*, New Delhi: NCERT, 1994. (Article in a Journal)

Selltiz, Claire et. al. (1959), *Research Methods in Social Relations*, Rinehart & Winston, Holt, New York.

Dhanarajan, Gajaraj, "Access to Learning and Asian Open Universities: In Context"in the 12th Annual Conference of Asian Association of Open Universities, (1998) "*The Distance Learner*" The Open University of Hong Kong, Hong Kong SAR, China, 4-6 Nov., (Conference Paper).

You would notice the following:

- All studies are arranged in alphabetical order
- The names of the authors are recorded by title and initials (not full name).
- To indicate two or three authors, 'and' is used between the first and the second, ',' between first and 'and' between second and third author.
- In case of more than three authors, only the name of the first author is mentioned followed by *et al*. (*et allibi*) or others.
- In case of a chapter in a book, after the author and chapter title and the name of the author or editor of the book.
- Titles of printed books, names of journals are highlighted by using 'italics' or by underlining (in case of manually typed material).
- Place of publication of a book precedes the name of the publisher separated by a ':'(colon).
- Names of journals are following by the relevant volume and issue numbers usually in the form 10(3) –Volume 10, Number 2 and page numbers.
- Unpublished thesis or dissertation titles are not highlighted and the word 'unpublished' is mentioned.

Referring Web-Based Documents

Computers have brought revolution in all sectors of development including education. Computers were conventionally used for data storage, processing and retrieval. Through internet, information can be accessed from any part of the world. As researchers, reviewing the relevant literature related to the problem under study is almost magnum opus. These days internet is a rich academic and professional resource. World Wide Web (WWW) is the easiest and most popularly used browsing mechanism on the Internet. Here, we will very briefly explain as how to write the references when we quote from any Web Site.

Citing E-Mail

E-Mail communications should be citied as personal communications as noted in APA's publication Manual http://www.apa.org/journals/webref.html. Personal Communications are hot cited in the reference list. The format in the text should be as:

Citing a Web Site

When you access the entire Web site (not a specific document on the site), you just give the address of the site in the text. It is not necessary to enter in the reference section.

For example,

http://www.ignou.ac.in (IGNOU's website)

http://www.webct.com/ (This site provides tools for development of web-based courses)

Citation of specific document on a web site has a similar format to that for print. Here, we give few examples of how to cite documents. The Web information is given at the end of the reference section. The date of retrieval of the site should be given because documents on the Web can change in content or they may be removed from a site.

Example

Duchier, D. (1996), Hypertext, New York: Intelligent Software Group. [Online] http://www.isg.sfu.ca/duchier/misc/hypertext - review/chapter4.htm]Accessed on 25/1/99].

Flinn, S. (1996), Exploiting information structure to guide visual browsing and exploratory search in distributed information systems [Online] http://www.cs.ubc.ca/reading-room/[Accessed June 1998]

If you have to cite some specific parts of a web document, indicate the chapter, figure, table as required.

Appendices

Usually, the appendices present the raw data, the true copy of the tools used in the study, important statistical calculations, photographs and charts not used inside the text. These are ordered serially like Appendix-1, Appendix-2, or they can be serialised with capital letters (Appendix A, Appendix B) etc. to facilitate referencing within the text. The appendices provide reference facilities to readers and others interested in that particular field of investigation.

Activity

1) Take any report and check whether the references are written in the standard form. If not, try to rewrite them properly.

.....

.....

.....

.....

.....

2) Examine the appendices in the same report. Are all of them essential for the report? Comment.

.....

.....

.....

.....

.....

.....

4.6 LET US SUM UP

In this chapter, we focused on research reporting as a professional activity. The purpose of writing the report depends on the reason behind undertaking the research study. It could be for obtaining a degree, or as a project report to be submitted to the funding agency, etc. Once submitted, the funding agency and the educational managers could utilise the findings and recommendations to achieve their objectives; other researchers may seek guidance from it and lastly, the findings may be used for developing new theories in the discipline concerned.

A research report has three parts: the beginning, the main body and the end. The beginning includes: cover or the title page, acknowledgements, table of contents, the list of tables and the list of figures. The main body normally contains an introduction, review of the relevant literature, objectives, hypotheses, research design (research methodology, population and sample, tools, procedure of collecting data), analysis and interpretation of data, the main findings and conclusion (that also includes its educational implications and suggestions for further studies). While discussing the main body, we have talked about the style of writing the report, style and placement of footnotes and reference, the typing process and the format and placement of tables and figures. We close the discussion with notes on the style, arrangement and placement of references and appendices which constitute the end of a research report.

4.7 FURTHER READINGS AND REFERENCES

Baker, L., Therese (1988), *Doing Social Research,* McGraw Hill, New York.

Black, James A. and Champion, Dean J. (1976), *Methods and Issues in Social Research,* John Wiley, New York.

Bryman, A. Social Research Methods. Oxford University Press (2012). India.

Kerlinger, Fred R. (1964), *Foundations of Behavioral Research,* Surjeet Publications, Delhi.

Kidder, Louise H. (1981), *Research Methods in Social Relations,* Holt, New York.

Lal Das, D.K. (2000), *Practice of Social Research : A Social Work Perspective,* Rawat Publications, Jaipur.

Monette, Duane R. et. al. (1986), *Applied Social Research: Tool For the Human Services,* Holt, Chicago.

Nachmias D and Nachmias C. (1981), *Research Methods in the Social Sciences*, St. Martins Press, New York.

Rubin, Allen & Babbie E. (1989), *Research Methodology for Social Work*, Belmont, Wadsworth, California.