UNIT 3 BASICS OF STATISTICAL TECHNIQUES

*Prof. Sushma Batra & Prof. Archana Kaushik

Contents

3.0 Objectives

3.1 Introduction

3.2 Statistical Methods: Functions

3.3 Measures of Central Tendencies

3.5 Measures of Dispersion

3.6 Let Us Sum Up

3.7 Key Words

3.8 Suggested Readings

3.0 OBJECTIVES

In this unit, certain basic statistical techniques used in social research have been discussed. After reading this unit, you should be able to:

- calculate mean, median and mode,
- differentiate between discrete and continuous series of data and application of statistical formulae accordingly,
- understand applicability of measures of dispersion; and
- develop insight into use of statistics for data interpretation and analysis.

3.1 INTRODUCTION

Numerical data collected in research studies can be analyzed quantitatively using statistical tools in two different ways - descriptive statistics and inferential statistics. Descriptive analysis refers to statistically describing, aggregating, and presenting the constructs of interest or associations between these constructs. Inferential analysis refers to the statistical testing of hypotheses (theory testing). In this unit, basic statistical techniques used for descriptive analysis are given and briefly inferential analysis is mentioned. As mostly researchers rely on computer software like SPSS for data analysis and interpretation, a rudimentary familiarization with statistical techniques would go a long way in ensuring which technique is to be used in which type of data. Failing in this basic understanding would not only jeopardize the entire research efforts but also make the researcher confused and caught up amidst huge data.

As discussed in earlier units, Uni-variate analysis, or analysis of a single variable, refers to a set of statistical techniques that can describe the general properties of one variable. Uni-variate statistics include: (1) frequency distribution, (2)

*Prof. Sushma Batra & Prof. Archana Kaushik, Department of Social Work, University of Delhi

central tendency, and (3) dispersion. The frequency distribution of a variable is a summary of the frequency (or percentages) of individual values or ranges of values for that variable.

Bi-variate analysis examines how two variables are related to each other. The most common bi-variate statistic is the bi-variate correlation (often, simply called 'correlation'), which is a number between -1 and +1 denoting the strength of the relationship between two variables.

3.2 STATISTICAL METHODS: FUNCTIONS

The knowledge of statistics helps the social workers in arriving at inferences as per the research objectives. These statistical procedures are largely becoming a part of all the social science researchers. They enhance the effectiveness and the efficiency of the services provided by the professional social workers. Statistics is, thus, a branch of applied mathematics and helps the researcher to understand the complex social phenomena better and lend precision to his/her data. It is this field of mathematics which is the subject matter of this unit.

Definition

The most comprehensive definition of statistics has been given by Prof. Horace Secrist who defined statistics as aggregates of facts affected to marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other.

This definition clearly points out certain characteristics which numerical data must possess in order to be called statistics. These include:

- Statistics are aggregates of facts.
- Statistics are affected to a marked extent by multiplicity of causes.
- Statistics are enumerated or estimated according to reasonable standard of accuracy.
- Statistics are numerically expressed.
- Statistics are collected in a systematic manner.
- Statistics are collected for a predetermined purpose.
- Statistics should be placed in relation to each other.

Statistical Methods

The large volume of numerical information gives rise to the need for systematic methods which can be used to organize, present, analyze and interpret the information effectively. Statistical methods are primarily developed to meet this need. Croxton and Cowden have given a very simple and concise definition of statistics. In their view 'Statistics may be defined as the collection, presentation, analysis and interpretation of numerical data. This definition clearly points out four stages in a statistical investigation namely: collection, presentation, analysis and interpretation of data. However, to the above stages one more stage may be added and that is the organization of data. Thus, statistics may be defined as the science of collection, organization, presentation, analysis and interpretation of numerical data.

Functions of Statistics

The following are the important functions of the science of statistics:

- It presents facts in a definite form.
- It simplifies mass of figures.
- It facilitates comparison.
- It helps in formulating and testing of hypothesis.
- It helps in making predictions.
- It helps in the formulation of suitable policies.

Statistics and Computers

The development of statistics has been closely related to the evolution of electronic computers as it is possible to perform millions of calculations in mere seconds with the help of computers. In spite of the fact that it is possible to do all the calculations with the computer, statistics does not lose its importance as it is possible to draw inferences only if the researcher has comprehensive knowledge of what to do with the data and which in turn is possible only if the researcher has the knowledge of statistics. It enables the researcher in making sense out of the available data. The knowledge of statistics helps the researcher in taking decisions regarding applicability of various tests with the help of computer. Therefore, while analyzing the data the importance of statistics cannot be underestimated.

In statistics we need to learn about the measures dealing with one variable, two variables and more than two variables. The basic measures which summarize the data into one figure are the measures of central tendency and dispersion. The measures used to determine relationship between two or more than two variables are called measures of correlation. The description in this chapter is restricted to measures of central tendency and dispersion.

Measures of central tendency describe how the data cluster together around a central point. There are three main measures of central tendency: the mean, the median and the mode. The measures of dispersion commonly used are range, quartile deviation, mean deviation and the standard deviation.

Check Your Progress I

Note: Use the space provided for your answer.

1) What are the functions of statistics?

....

....

....

....

3.3 MEASURES OF CENTRAL TENDENCIES

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 simple average.

Ungrouped Data

The data collected for the purpose of a statistical inquiry are simple figures without any form or structure. Data obtained in this way are in a raw state for they have not gone through any statistical treatment. This shapeless mass of data is known as ungrouped data or raw data. Consider the data presented in Table 3.1

Table 3.1: Marks in Social Research obtained by 20 students.

Roll No.	Marks	Roll No.	Marks
1	15	11	15
2	17	12	18
3	18	13	7
4	10	14	8
5	13	15	9
6	11	16	20
7	8	17	13
8	12	18	11
9	13	19	12
10	19	20	13

Ungrouped data presented in this manner are not capable of being rapidly or easily interpreted.

Grouped Data

In order that data may be more readily comprehensible, grouping further reduces the bulk of these data. A first step in such a grouping would be achieved by representing the repetitions of a particular mark by tallies instead of corresponding to any given marks in the frequency of that class (usually denoted by the letter 'f') as shown in table 3.2

Table 3.2: Tally Sheet showing the marks obtained by 20 students

Marks	Tally	Frequency
7	1	1
8	11	2
9	1	1
10	1	1
11	11	2
12	111	3

13	111	3
15	11	2
17	1	1
18	11	2
19	1	1
20	1	1
Total	**20**	**20**

Discrete and Continuous Data

Data may be either discrete or continuous. A discrete series is formed from data which are capable of exact measurement. In other words, certain kinds of data are discrete in that not all values are possible. A family may have exactly 1, 2, 3 or even 4 children but it cannot have an integer. It has to be a whole number which is capable of being counted. Such data would give rise to discrete data (Table 3.3)

Table 3.3: Frequency Distribution: Discrete data

No. of children/family	No. of couples
0	6
1	7
2	11
3	15
4	1
Total	**30**

Certain other kinds of data are continuous in that all values are possible. That is, there are certain items which are not capable of exact measurement like weight or height. We cannot count the number of students whose height are exactly 160 cm. In such cases the data are given in relation to groups or class intervals. For example, we can count the number of students whose height are between 155 cm and 165cm. These types of data are called continuous data (Table 3.4)

Table 3.4: Frequency Distribution: Continuous Series

Height in Cm.	No of students frequency (f)
150-155	12
156-160	15
161-165	18
166-170	17
171-175	15
176-180	14
Total	**91**

Class Intervals

Each group of five consecutive values of heights, namely, 150-155, 155-160, 160-165, etc., is called class. Since each class includes five values, 05 is the magnitude or width of the class, commonly known as class interval. The first figure of each class is called its lower limit, and the last figure of each class is known as the upper limit. Every class interval has a mid- point which is mid-way between the upper and lower limits. Table 3.5 illustrate the lower limits, upper limit and mid point of the classes, taking the above example into consideration.

Table 3.5: Lower limits, Mid Points and Upper limits in continuous data

Intervals	Mid point	Lower limit	Upper limit
150-155	152.5 *	150	155
156-160	157.5	155	160
161-165	162.5	160	165

* **E.g.** $\frac{150+155}{2} = 152.5$

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

The Mean

The mean is 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 mean is defined as the sum of the values of all the items and dividing the total by the number of items. An example will help us learn how to calculate the arithmetic mean. Let us suppose that eight students receive 54, 58, 60, 62, 70, 72, 75 and 77 marks respectively, in an examination, the mean of marks will be:

$$\text{Mean} = \frac{54+58+60+62+70+72+75+77}{8} = \frac{528}{8} = 66$$

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.

Illustration

Table 3.6: Distribution of Rag Pickers by their Daily Income

S. No.	Daily Income (in Rs.)	Number of Rag-Pickers
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**

Solution:

Daily Income (in Rs.)	Mid-values (m)	Number Rag-Pickers (f)	m x f
110-130	120*	15	1800
130-150	140	30	4200
150-170	160	60	9600
170-190	180	95	17100
190-210	200	82	16400
210-230	220	75	16500
230-250	240	23	5520
		$\sum f = N = 380$	$\sum mf = 71120$

Calculation:

$$\text{*Mid-value} = \frac{\text{Lower limit + Upper limit}}{2} = \frac{110 + 130}{2}$$

$$= \frac{240}{2} = 120$$

$$\overline{X} = \frac{\sum mf}{\sum f} = \frac{\sum mf}{N}$$

$$= \frac{71120}{380} = 187.16$$

Mean = Rs.187.16 (approximately)

Solution:

Table 3.7

Monthly wages (in Rs.)	Mid-values (m)	No. of workers (f)	Deviation from assumed mean. 180 (dx)	Step deviation (d)	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

$$\text{Mean } (X) = a + \frac{\Sigma fd}{N} \times i$$

Where 'a' stands for the assumed mean, Σfd 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 :

$$= \quad 180 + \frac{136}{380} \text{ x } 20$$

$$= \quad 180 + 7.16$$

Mean = Rs. 187.16 (approximately)

Merits

Arithmetic mean is most widely used in practice because :

- It is simplest average to understand
- It is easy to compute.
- Value is rigid
- Takes into consideration all the items.
- Value is reliable- sampling stability

Limitations

- Since the value of mean depends on each and every item of the series, extreme items- very small and very large unduly effect the valued average.
- Mean cannot be computed in open end classes, we need to go by assumption.
- It is a good mean only when population follows a normal distribution.

The Median

The median is another simple measure of central tendency. We sometimes want to locate the position of the middle item when data have been arranged. This measure is also known as positional averages. We define the median as the size of the middle item when the items are arrayed 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.

Table 3.8: Daily Income of Rag-Pickers

Daily Income in Rs.	Number of Rag-pickers (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 3.8 above the median value is that value on either side of which $\frac{N}{2}$ or $\frac{380}{2}$ or 190th items 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).

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 x 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:

$$X = L + \frac{N/2 - C}{f} \times i$$

Where

X = median,

L = lower limit of the class in which median lies

N = total number of items

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

'f = frequency of the median class

i = class interval of the median class.

$$= 170 + \frac{\frac{380}{2} - 105}{95} \times (190 - 170)$$

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

$$= 170 + \frac{85}{95} \times (190 - 170)$$

= 170 + (0.8947 x 20)

= 187.89 (approximately)

Median Income = Rs.187.89 (approximately)

Merits & Limitations of Median

Merits

- It is especially useful in case of open end classes since only the position and not the value of items must be known.
- It is also useful in unequal class intervals.
- Extreme items do not affect the median as they do to the mean.
- Skewed distribution where Arithmetic mean would be distorted by extreme values, the median is especially useful.
- Most appropriate average in dealing with qualitative data.
- The value of median can be determined graphically whereas mean cannot be determined graphically.
- It gives us the middle value of the distribution which is mostly required.

Limitations

- For calculating median, it is necessary to arrange the data.
- Since it is a positioned average, its value is not determined by each and every observation.

- Not capable of further statistical treatment.
- The value of median is affected more by sampling fluctuations, than the value of arithmetic mean.
- The median in some cases cannot be computed exactly as the mean. This is true when number of items included in the series is even and that too in discrete series.

The 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 occurs 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

The first step in calculation of mode is to find out the point of maximum concentration with the help of grouping method. The procedure of grouping 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 the process of grouping locates the class of maximum concentration the value of mode is interpolated by the use of the following formula.

$$\text{Mode (X)} = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Where X stands for the mode, L is the lower limit of the modal class, f_0 stands for the frequencies of the preceding class, f1 stands for the frequencies of the modal class, f_1 for the frequencies of the succeeding class and i stands for the class interval of the modal class.

Illustration:

Table 3.9: Weekly Family Income (in Rs.)

Weekly Income	Number of families
100 – 200	5
200 – 300	$6 = f_0$
300 – 400	$15 = f_1$
400 – 500	$10 = f_{2,}$
500 – 600	5
600 – 700	4
700 – 800	3
800 – 900	2
Total	**N = 50**

Table 3.10: Location of Modal Class by Grouping

<table>
<tr><th>Weekly Income</th><th>F(1)</th><th>(2)</th><th>(3)</th><th>(4)</th><th>(5)</th><th>(6)</th></tr>
<tr><td>100 – 200</td><td>5</td><td rowspan="2">11</td><td></td><td rowspan="3">26</td><td></td><td></td></tr>
<tr><td>200 – 300</td><td>6</td><td rowspan="2">21</td><td rowspan="3">31</td><td></td></tr>
<tr><td>300 – 400</td><td>15</td><td rowspan="2">25</td><td rowspan="3">30</td></tr>
<tr><td>400 – 500</td><td>10</td><td rowspan="2">15</td><td rowspan="3">19</td></tr>
<tr><td>500 – 600</td><td>5</td><td rowspan="2">9</td><td rowspan="3">12</td></tr>
<tr><td>600 – 700</td><td>4</td><td rowspan="2">7</td><td rowspan="3">9</td></tr>
<tr><td>700 – 800</td><td>3</td><td rowspan="2">5</td><td></td></tr>
<tr><td>800 – 900</td><td>2</td><td></td><td></td><td></td></tr>
</table>

Table 3.11: 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	1	3	(6)	3	1			

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

$$X = L + \frac{f_1 - f_{0.}}{2f_1 - f_0 - f_2} \times i$$

$$\bar{X} = 300 + \frac{15 - 6}{2 \times 15 - 6 - 10} \times 100$$

$$= 300 + \frac{9}{14} \times 100$$

= 300 + 64.29

= 364.29 (approximately)

Merits & Limi♠tations of Mode

Merits

- It is the most frequently occurring value- hence in greater demand.
- Mode is not unduly affected by extreme values.
- Can be used in open-end distributions
- Mostly used to describe qualitative phenomenon eg. Consumer preference of different products of daily use.
- Value can be determined graphically.

Limitations

- The value of mode cannot always be determined eg. Bimodal series.
- Not capable of further statistical treatment.
- The value of mode is not based on each and every item of the series.
- It is not a rigidly defined measure with the use of different formulas we get different results.
- Not much in use.

Usefulness of Mode

1. Used when most typical value of a distribution is desired.
2. Especially useful in skewed distribution.

Relationship between Mean, Median and Mode

Mode = 3 Median – 2 Mean

Check Your Progress II

Note: Use the space provided for your answer.

1) Define the terms measures of central tendency.

....

....

....

....

....

3.4 MEASURES OF DISPERSION

In social work research, we often wish to know the extent of homogeneity and heterogeneity among respondents with respect to a given characteristic. Any set of social data is values, which are heterogeneous. The set of social data is characterized by the heterogeneity of values. In fact, the extent to which they are heterogeneous or vary among themselves is of basic 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.

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. 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 is 850 the range will be 700.

Semi-Inter-Quartile Range or Quartile Deviation

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, where Q_1 and Q_3 stand for first and the third quartiles, the semi-interquartile range or quartile deviation.

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

Calculation of Quartile Deviation (QD)

Table 3.12: Weekly Family Income in (Rs.)

Weekly Income	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 3.13

S. No. (1)	Weekly Income (in Rs.) (2)	Number of Families (3)	Cumulative Frequency (CF) (4)
1	100 – 200	5	5
2	200 – 300	6	11 = c
3	Q_1 – 300 – 400	15 = f	26
4	400 – 500	10	36 = c
5	Q_3 - 500 – 600	5 = f	41
6	600 – 700	4	45
7	700 – 800	3	48
8	800 – 900	2	50
	Total	**N = 50**	

$$Q_1 = 1_1 + \frac{-C}{F} (I)$$

$$= 300 + \frac{12.5 - 11}{15} \times 100$$

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

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

$$= 300 + 10$$

$$= 310$$

$$Q_3 = L_1 + \frac{-C}{5} (I)$$

$$= 500 + \frac{37.5 - 36}{5} \times 100$$

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

$= 500 + (0.3 \text{ x } 100)$

$= 500 + 30$

$= 530$

$= Q_3 – Q_1$

$= 530 – 310 = 220$

$$QD = \frac{Q_3 - Q_1}{2}$$

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

$= 110$

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.

$$\text{Symbolically, coefficient of Q.D.} = \frac{\frac{.Q^3 - Q^1}{2}}{\frac{Q_3 + Q_1}{2}} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Applying this to the preceding illustration we get,

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

Mean Deviation

Quartile deviation suffers from a serious drawback; they are calculated by taking into consideration only two values of a series. As a result, the composition of the series is entirely ignored. To avoid this defect, 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 Mean Deviation.

Illustration:

Table 3.14: Weekly Family Income (in Rs.)

Weekly Income	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**

Solution :

Table 3.15: Weekly Family Income (in Rs.)

	Weekly Income	Mid Value	No. of families (f)	Cumulative frequency	Deviation from median 400 | d |	'f | d |
	100-200	150	5	5	250	1250
	200-300	250	5	10=C	150	750
Median Group	300-400	350	15 = f	25	50	750
	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	Procedure	Application to Table 3.15
1	Calculate the median of the distribution	$X = L + \dfrac{\dfrac{N}{2} - C.}{'f} \times 'I$ $= 300 + \dfrac{\dfrac{50}{2} - 10}{15} \times 100$ $= 300 + \dfrac{25 - 10}{15} \times 100$ $= 300 + \dfrac{15}{15} \times 100$ $= 300 + (1 \times 100)$ $= 300 + 100 = 400$
2	Find mid-points of each class	$= \dfrac{100+200}{2} = \dfrac{300}{2} = 150,....$
3	Find absolute deviation – |d| of each mid – points from median (400)	$\| 150 - 400 \| = \| -250 \|$ $= 250,.....$
4	Find total absolute deviation by multiplying the frequency of each class by the deviation of its mid – points from the median (f| d |)	$5 \times 250 = 1250,.....$
5	Find the sum of products of frequency and deviations (f |d |)	F | d | = 7400
6	Compute Mean Deviation	$1\ (X) = \dfrac{f \| d \|}{N} = \dfrac{7400}{50} = 148$

Standard Deviation

The most useful and frequently used measure of dispersian 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.

$$= \frac{d^2}{N}$$

Where a (Greek letter sigma) stands for the standard deviation, Σd2 for the sum of the squares of the deviation measured from mean and N for the number of items.

$$\alpha = \frac{\Sigma \ d^2}{N}$$

Where a (Greek letter sigma) stands for the standard deviation, Σd^2 for the sum of the squares of the deviation measured from mean and N for the number of items.

Calculation of Standard Deviation

In a continuous series the class intervals are represented by their midpoints. However, usually the class-intervals are of equal size and thus, the deviations from the assumed average is expressed in class interval units. Alternatively, step deviation is found out by dividing the deviations by the magnitude of the class interval. Thus, the formula for computing standard deviation is written as follows;

$$\alpha = \frac{\sum fd^2}{N} - \frac{\sum f d^{\ 2}}{N} \times t$$

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

Table 3.16: Weekly Family Income (in Rs.)

Sl No.	Weekly Income	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 3.17

S. No.	Weekly Income	Mid values (m)	Number of families (f)	Step deviation from ass. Ave (450) (d)	f d	d^2	Fd_1
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		$\Sigma fd^2 = 164$	

Step	Procedure	Application to Table 3.17
1	Find the mid-points of the various classes	$\frac{100+200}{2} = \frac{300}{2} = 150$,....
2	Assume a mid-points 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 stepdeviation (d)	(1) 150–450 =–300/3= –3…
4	The deviations are multiplied by the frequency of each class (fd)	(–3) (5) = –15 (–2) (6) = –1
5	Find the aggregate of products of step 4 (Σ fd)	Σ fd = –12
6	Square the deviations (d^2)	(–3) (–3) = 9,...
7	Squared deviations are multiplied by the respective frequencies (fd^2)	9 × 5 = 45,...
8	Find the aggregate of products of step 7 (Σfd^2)	$\Sigma fd^2 = 164$
9	Compute standard deviation with the help of the formula	$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times$ $\sigma = \sqrt{\frac{164}{50} - \left(\frac{-12}{50}\right)^2} \times$ $= \sqrt{3.28 - 0.0576 \times 100}$

		$= \sqrt{3.2224 \times 100}$ $= 1.795 \times 100$ = 179.51 (approximately)

Check Your Progress III

Note: Use the space provided for your answer.

1) Define the term measures of dispersion.

....

....

....

....

....

3.5 LET US SUM UP

Knowledge of statistics helps the social worker in two ways. First, the knowledge allows the social worker to be able to analyse the data and draw inferences. Second, as a consumer of researches it enables him/her to understand the analysis of data used in research reports.

Ungrouped data are not capable of being rapidly or easily interpreted. In order that data may be more readily comprehensible data can be grouped.

Mean, median and mode are the three measures of central tendency. Mean is the arithmetic average of a distribution. It is computed by dividing the sum of all values of observations by the total number of values. Median is a point in an array, which divides a data set into two equal halves in such a way that all the values in one half will be greater than the median value and all the values in other half will be smaller than the median value. Mode is a most frequently occurring value in a distribution.

The range, and standard deviation are the most commonly used measures of variability. The range is the difference between the two extreme values. The square root of the average of the squared deviations of the measures or values from their mean is known as standard deviation.

3.6 KEY WORDS

Central Tendency : The typical value in any data array that represents the middle position.

Descriptive Statistics : Statistical methods used to illustrate the meaning of the data in a study without drawing any inferences.

Frequency : The number of observations for a variable.

Mean	:	Another word for average; in a distribution of ordinal or scale values, the sum of scale values divided by the number of values being considered.
Median	:	In a distribution of ordinal or scale values, the exact mid-point so that 50% of the values fall higher and 50% of values fall lower in the distribution.
Mode	:	In a distribution of nominal, ordinal or scale values , the most commonly occuring value.
Ungrouped Data	:	Data in the first of simple figures without any of nominal, ordinal or scale values , the most commonly occuring value.
Continous Series	:	We have record of the frequenies with which the observation appear in each of the class-intervals.
Range	:	The difference between two extreme values.
Semi-Inter-Quartile Range	:	The Difference between the values of first and the third quartiles.

3.7 SUGGESTED READINGS

Faherty, Vincent E. (2008). Compassionate Statistics: Applied Quantitative Analysis for Social Services. London: Sage

Blalock Jr., H.M. (1960). Social Statistics. New York: McGraw Hill Book Company, Inc.

Nicola, B., Richard, K. & Rose Mary, S.(2003). SPSS for Psychologists: A Guide to Data Analysis Using SPSS for Windows, Palgrave Macmillan.

Montcalm, D. & Royse, D. (2002). Data Analysis for Social Workers. London: Allyn and Bacon

Laldas D.K. (2000). Practice of Social Research: Social Work Perspective. New Delhi: Rawat Publications

Elhance, D.N. (1984) *Fundamental of Statistics*, Kitab mahal, Allahbad.

Freud, J.E. (1977), *Modern Elemenatary Statistics*, Prentice hall, New Delhi.

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

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

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

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