# BLOCK 2

# MEASURES OF CENTRAL TENDENCY AND VARIABILITY

# BLOCK INTRODUCTION

The second block of this course is about measures of central tendency and measures of variability. This block has three units, unit 3, 4 and 5. **Unit three** will mainly focus on the concept of central tendency of data and the different measures of central tendency, namely, mean, median and mode. In this unit we will not only discuss about the properties, advantages and limitations of mean, median and mode but also how to compute them. The **fourth unit** will deal with measures of variability. In this unit, we will discuss about the concept of variability in data with a focus on the functions of variability, as well as, on absolute dispersion and relative dispersion. The unit will also focus on the different measures of variability, namely, range, quartile deviation, standard deviation, average deviation or mean deviation, standard deviation and variance. The **fifth unit** in this block covers computation of various measures of variability.

# UNIT 3 INTRODUCTION TO MEASURES OF CENTRAL TENDENCY*

**Structure**




* Prof. Suhas Shetgovekar, Faculty, Discipline of Psychology, School of Social Sciences, IGNOU, New Delhi

## 3.0 OBJECTIVES

After reading this unit, you will be able to:

- explain the concept of central tendency of data;
- describe the different measures of central tendency;
- discuss the properties, advantages and limitations of mean, median and mode; and
- compute measures of central tendency for ungrouped and grouped data.

## 3.1 INTRODUCTION

Suppose you have data, for instance, marks in psychology obtained by students in 12th standard and you want to analyse it statistically, what statistical techniques will you employ? You can of course organise the data with the help of classification and tabulation that we discussed in the previous Unit and the data can also be graphically represented. But if you want to further analyse the data then you can compute the average marks obtained by the whole class or find the midpoint for marks above and below which will lie half of the students or you can also find out most frequent marks obtained by the students. The techniques you are employing here are mean, median and mode. These are called measures of central tendency and can be categorised under descriptive statistics.

In the previous unit, we discussed about classification, tabulation and also graphical representations of data. In the present unit, we will discuss the measures of central tendency, viz., mean, median and mode. We will not only understand what these techniques are, but will also focus on their properties, advantages and limitations. Further, we will also learn how to compute mean, median and mode for grouped and ungrouped data.

## 3.2 CONCEPT OF CENTRAL TENDENCY OF DATA

Measures of central tendency provides a single value that indicates the general magnitude of the data and this single value provides information about the characteristics of the data by identifying the value at or near the central location of the data (Bordens and Abbott, 2011). King and Minium (2013) described measures of central tendency as a summary figure that helps in describing a central location for a certain group of scores. Tate (1955, page 78) defined measures of central tendency as "a sort of average or typical value of the items in the series and its function is to summarise the series in terms of this average value".

The main functions of measures of central tendency are as follows:

1) They provide a summary figure with the help of which the central location of the whole data can be explained. When we compute an average of a certain group we get an idea about the whole data.

2) Large amount of data can be easily reduced to a single figure. Mean, median and mode can be computed for a large data and a single figure can be derived.

3) When mean is computed for a certain sample, it will help gauge the population mean.

4) The results obtained from computing measures of central tendency will help in making certain decisions. This holds true not only to decisions with regard to research but could have applications in varied areas like policy making, marketing and sales and so on.

5) Comparison can be carried out based on single figures computed with the help of measures of central tendency. For example, with regard to performance of students in mathematics test, the mean marks obtained by girls and the mean marks obtained by boys can be compared.

A good measure of central tendency needs to have the following characteristics:

1) The definition of the central tendency needs to be adequately specified and should be clear. It should not be subject to varied interpretations and needs to be unaffected by any individual bias. The definition should be rigid so that a stable value is obtained that represents the data.

2) The measure of central tendency should be easy to understand and easy to compute. It should not involve elaborate mathematical calculations.

3) For the value obtained from the computation of measures of central tendency to be representative of the data, the whole data needs to be computed.

4) The data needs to be collected from a sample that truly represents the population. The sample thus needs to be randomly selected.

5) The measure of central tendency needs to display sampling stability and should not be affected by any fluctuations in the sample. For example, if two different researchers obtain a representative sample from a same population, the means computed by them for their respective sample should display least variation.

6) The measure of central tendency should not be affected by outliers. Outliers are extreme values in data or distribution.

7) The measure of central tendency should render itself to further mathematical computations.

**Check Your Progress I**

1) Define measures of central tendency.

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

2) List the functions of measures of central tendency.

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

## 3.3 DIFFERENT MEASURES OF CENTRAL TENDENCY

As the concept of central tendency is now clear, we will now proceed to discuss the three measures of central tendency. The three measures of central tendency that we will be discussing are:

1) Mean or Arithmetic mean
2) Median
3) Mode

In this section of the Unit, we will try to understand these concepts and then in the next section we will be focusing on the properties, advantages and limitations of each of these measures.

### 3.3.1 Mean or Arithmetic Mean

Mean for sample is denoted by symbol 'M or $\overline{x}$ ('x-bar')' and mean for population is denoted by '$\mu$' (mu). It is one of the most commonly used measures of central tendency and is often referred to as average. It can also be termed as one of the most sensitive measure of central tendency as all the scores in a data are taken in to consideration when it is computed (Bordens and Abbott, 2011). Further statistical techniques can be computed based on mean, thus, making it even more useful.

Mean is a total of all the scores in data divided by the total number of scores. For example, if there are 100 students in a class and we want to find mean or average marks obtained by them in a psychology test, we will add all their marks and divide by 100, (that is the number of students) to obtain mean.

### 3.3.2 Median

Median is a point in any distribution below and above which lie half of the scores. Median is also referred to as $P_{50}$ (King and Minium, 2008). The symbol for median is '$M_d$'. As stated by Bordens and Abbott (2011, page 411), 'median is the middle score in an ordered distribution'. If we take the example discussed earlier of the marks obtained by 100 students in a psychology test, these marks are to be arranged in an order, either ascending or descending. The middle score in this distribution is then identified as median. Though this would seem easy for an odd number of scores, in case of even number of scores a certain procedure is followed that will be discussed when we learn how to compute median later in this unit.

### 3.3.3 Mode

Mode is denoted by symbol '$M_o$'. Mode is the score in a distribution that occurs most frequently. Taking the example of the marks obtained by a group of 100 students in psychology test discussed earlier, if out of these 100 students, 10 students obtained 35 marks. 35 is thus, most frequently occurring value and will be termed as mode. Certain distributions can be bimodal as well, where there are two modes. For instance if there were other 10 students in this group of 100 students, who secured 47 marks, 47 is the value that is occurring as frequently as 35 and thus, will be termed as mode along with 35. In a similar way, when there are three modes, the term used is trimodal and when there are four or more modes, we use the term multimodal.

Though if the scores in a distribution greatly vary then it is possible that there is no mode. Mode as such does not provide an adequate characterisation of the distribution because it just takes in to consideration the most frequent scores and other scores are not considered.

**How to choose a measure of central tendency?**

The choice of a measure of central tendency will depend on first of all, the scales of measurement that we discussed in the first unit. For nominal scales one can compute mode but not mean or median. For example, in case of males and females, the males can be coded as 1 and females can be coded as 2 (or vice versa) in such a case, we can compute frequently occurring score, that will provide us information whether there are more males or more females. However, it is not possible to compute mean or median. With regard to ordinal scale, median or mode can be used. For example, if we rank the students basedon their performance in mathematics test, it is possible to find median belowand above which lie half of the ranks. Mode can also be computed if morethan one student gets same rank. With regard to interval scale and ratio scalemean can be computed.

Yet another aspect that is important while making a choice with regard to which measure of central tendency to use is, whether the data is normally distributed or not. If the data is normally distributed we will compute mean and if it is not normally distributed, we will compute median or mode. This is because mean may not adequately represent the data when the data is not normally distributed. We will discuss normal distribution in detail in the last unit (unit 8) of this course.

**Check Your Progress II**

1) Describe mean, median and mode.

| Measure | Description | Example |
|---|---|---|
| **Mean** | | |

| | | |
|---|---|---|
| **Median** | | |
| **Mode** | | |

2) How to choose a measure of central tendency?

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

## 3.4 PROPERTIES, ADVANTAGES AND LIMITATIONS OF MEAN, MEDIAN AND MODE

Let us now discuss the properties, advantages and limitations of mean, median and mode.

### 3.4.1 Properties of Mean

1) Mean is sensitive to the actual position of each and every score in a distribution and if another score is included in the distribution, then the mean or average of that distribution will change. For example, mean of the scores 5, 4 ,6 ,3, 2 is 4 [We got the value 4 by adding 5+4+6+3+2= 20 and then dividing it by 5, that is the total number of scores (N)]. But if we change the scores to 5, 4, 6, 3, 2, 8, the mean will be 4.67 [We got the value 4.67 by adding 5+4+6+3+2+8= 28 and then dividing it by 6, that is the total number of scores (N)]

2) Mean denotes a balance point of any distribution and the total of positive deviations from the mean is equal to the negative deviations from the mean (King and Minium, 2008).

3) Mean is especially effective when we want the measure of central tendency to reflect the sum of the scores.

### 3.4.2 Advantages of Mean

1) The definition of mean is rigid which is a quality of a good measure of central tendency.

2) It is not only easy to understand but also easy to calculate.

3) All the scores in the distribution are considered when mean is computed.

4) Further mathematical calculations can be carried out on the basis of mean.

5) Fluctuations in sampling are least likely to affect mean.

### 3.4.3 Limitations of Mean

1) Outliers or extreme values can have an impact on mean.

2) When there are open ended classes, such as 10 and above or below 5, mean cannot be computed. In such cases median and mode can be computed. This is mainly because in such distributions mid point cannot be determined to carry out calculations.

3) If a score in the data is missing or lost or not clear, then mean cannot be computed unless mean is computed for rest of the data by not considering the lost score and dropping it all together.

4) It is not possible to determine mean through inspection. Further, it cannot be determined based on a graph.

5) It is not suitable for data that is skewed or is very asymmetrical as then in such cases mean will not adequately represent the data.

### 3.4.4 Properties of Median

1) When compared to mean, median is less sensitive to extreme scores or outliers.

2) When a distribution is skewed or is asymmetrical median can be adequately used.

3) When a distribution is open ended, that is, actual score at one end of the distribution is not known, then median can be computed.

### 3.4.5 Advantages of Median

1) The definition of median is rigid which is a quality of a good measure of central tendency.

2) It is easy to understand and calculate.

3) It is not affected by outliers or extreme scores in data.

4) Unless the median falls in an open ended class, it can be computed for grouped data with open ended classes.

5) In certain cases it is possible to identify median through inspection as well as graphically.

### 3.4.6 Limitations of Median

1) Some statistical procedures using median are quite complex. Computation of median can be time consuming when large data is involved because the data needs to be arranged in an order before median is computed.

2) Median cannot be computed exactly when an ungrouped data is even. In such cases, median is estimated as mean of the scores in the middle of the distribution.

3) It is not based on each and every score in the distribution.

4) It can be affected by sampling fluctuations and thus can be termed as less stable than mean.

### 3.4.7 Properties of Mode

1) Mode can be used with variables that can be measured on nominal scale.

2) Mode is easier to compute than mean and media. But it is not used often because of lack of stability from one sample to another and also because a single set of data may possibly have more than one mode. Also, when there is more than one mode, then the modes cannot be termed to adequately measure central location.

3) Mode is not affected by outliers or extreme scores.

### 3.4.8 Advantages of Mode

1) It is not only easy to comprehend and calculate but it can also be determined by mere inspection.

2) It can be used with quantitative as well as qualitative data.

3) It is not affected by outliers or extreme scores.

4) Even if a distribution has one or more than one open ended classe(s), mode can easily be computed.

### 3.4. 9 Limitations of Mode

1) It is sometimes possible that the scores in the data vary from each other and in such cases the data may have no mode.

2) Mode cannot be rigidly defined.

3) In case of bimodal, trimodal or multimodal distribution, interpretation and comparison becomes difficult.

4) Mode is not based on the whole distribution.

5) It may not be possible to compute further mathematical procedures based on mode.

6) Sampling fluctuations can have an impact on mode.

**Check Your Progress II**

1) List the properties of mean.

.......................................................................................................................

.......................................................................................................................

.......................................................................................................................

.......................................................................................................................

.......................................................................................................................

2) List the advantages of median.

.......................................................................................................................

.......................................................................................................................

.......................................................................................................................

.......................................................................................................................

.......................................................................................................................

3) List the limitations of mode.

.......................................................................................................................

.......................................................................................................................

.......................................................................................................................

.......................................................................................................................

.......................................................................................................................

## 3.5 COMPUTATION OF MEASURES OF CENTRAL TENDENCY IN UNGROUPED AND GROUPED DATA

Now as we have developed a fair idea about the three measures of central tendency, we will move on to learn how to compute them. While computing each of these measures, we will do so for ungrouped and grouped data. Ungrouped and grouped data are explained as follows:

**Ungrouped data:** Any data that has not been categorised in any way is termed as an ungrouped data. For example, we have an individual who is 25 years old, another who is 30 years old and yet another individual who is 50 years old. These are independent figures and not organised in any way, thus they are ungrouped data.

**Grouped data:** A data that is categories or organised is termed as grouped data. Mainly such data is organised in frequency distribution. For example, we can have age range 26- 30 years, 31- 35 years, 36- 40 years and so on. Grouped data are convenient especially when the data is large.

### 3.5.1 Computation of Mean for Ungrouped Data

The formula for computing mean for ungrouped data is

$M = \Sigma X / N$

Where,

M = Mean

$\Sigma X$= Summation of scores in the distribution

N = Total number of scores.

Let us now compute mean with the help of an example

The scores obtained by 10 students on psychology test are as follows:

58 34 32 47 74 67 35 34 30 39

**Step 1**: In order to obtain mean for the above data we will first add the marks to obtain $\Sigma X$:

58+ 34+ 32+ 47+ 74+ 67+ 35+ 34+ 30+ 39 = 450

**Step 2**: Now using the formula, we will compute mean

$M = \Sigma X / N$

$\Sigma X = 450$ , $N = 10$ (Total number of students)

Thus,

$M = 450 / 10 = 45$

Thus, the mean obtained for the above data is 45

### 3.5.2 Computation of Mean for Grouped Data

The formula for computing mean for grouped data is

$M = \Sigma fX / N$

Where,

M= Mean

$\sum$= Summation

X= Midpoint of the distribution

$f$ = The respective frequency

N = Total number of scores.

Let us now compute mean with the help of an example.

A class of 30 students were given a psychology test and the marks obtained by them were categorised in to six categories. The lowest marks obtained were 10 and highest marks obtained were 35. A class interval of 5 was employed. The data is given as follows:

| Marks | Frequencies (*f*) | Midpoint (X) | *f*X |
|---|---|---|---|
| 35- 39 | 5 | 37 | 185 |
| 30-34 | 7 | 32 | 224 |
| 25-29 | 5 | 27 | 135 |
| 20-24 | 6 | 22 | 132 |
| 15-19 | 4 | 17 | 68 |
| 10- 14 | 3 | 12 | 36 |
| | **N= 30** | | **ΣfX = 780** |

The steps followed for computation of mean with grouped data are as follows:

**Step 1:** The data is arranged in a tabular form with marks grouped in categories with class interval of 5.

**Step 2:** Once the categories are created, the marks are entered under frequency column based on which category they fall under.

**Step 3:** The midpoints of the categories are computed and entered under X.

**Step 4:** *f*X is obtained by multiplying the frequencies and midpoints for each category.

**Step 5:** *f*X for all the categories are added to obtain ΣfX, in case of our example it is obtained as 780

**Step 6:** The formula $M = \Sigma fX / N$ is used, N is equal to 30.

$M = \Sigma fX / N$

$M = 780 / 30 = 26$

Thus, the mean obtained is 26

### 3.5.3 Computation of Mean by Shortcut Method (with Assumed mean)

In certain cases data is very large and it is not possible to compute each *f*X. In such situations, a short cut method with the help of assumed mean can be computed. A real mean can thus be computed with application of correction.

The formula is

$$\mathbf{M = AM + (\Sigma fx'/N \times i)}$$

Where,

AM= Assumed mean,

Σ = Summation

i = Class interval

$x' = \{(X - AM)/\ i\}$, X the midpoint of the scores in the interval

$f$ = the respective frequency of the midpoint

N = The total number of frequencies or students.

Let us discuss the steps followed for computation of mean with the help of an example given below:

| **Class Intervals (Marks)** | **Frequencies ($f$)** | **Midpoint (X)** | **$x' = \{(X - AM)/\ i\}$** | **$fx'$** |
|---|---|---|---|---|
| 35- 39 | 5 | 37 | 3 | 15 |
| 30-34 | 7 | 32 | 2 | 14 |
| 25-29 | 5 | 27 | 1 | 5 |
| 20-24 | 6 | 22 | 0 | 0 |
| 15-19 | 4 | 17 | -1 | -4 |
| 10- 14 | 3 | 12 | -2 | -6 |
| | **N= 30** | | | **$\Sigma fx'$= 24** |

**Step 1:** We will assume mean (AM) as 22.

**Step 2:** Difference is obtained between each of the midpoints and the assumed mean and then the same is divided by 'i' that is the class interval (5 in this case), these are then entered under column with heading $x' = \{(X - AM)/\ i\}$. The $x'$ for 22 will be 0.

**Step 3:** Frequency ($f$) is then multiplied with $x'$ to obtain $fx'$.

**Step 4:** All $fx'$ are added to obtain $\Sigma fx'$, in the present example it is 24.

**Step 5:** The formula for mean is now applied

**M =AM+($\Sigma$fx′/N×i)**

M= 22 + (24/30 x 5)

= 22 + 4 = 26

Thus, mean is obtained as 26.

And if you refer to the mean obtained by the direct method and mean obtained with the shortcut method, the mean is the same, that is 26.

### 3.5.4 Computation of Median for Ungrouped Data

With regard to computation of median for ungrouped data, different procedures are followed for data that is odd and data that is even.

**3.5.4.1 Odd Data:** When the data is odd the median is computed in the following manner:

Data: 58 34 32 47 74 67 35 34 30 (N= 9)

**Step 1**: First the data is to be arranged in either ascending or descending order. We will arrange the data in ascending order and it will look like this:

30 32 34 34 35 47 58 67 74

**Step 2**: The following formula is then used to compute Median:

**$M_d = (N+1)/2^{th}$ score**

Thus $(9+1)/2 = 10/2 = 5^{th}$ item

In our data the $5^{th}$ item is 35, that is the median of this data.

**3.5.4.2 Even data:** When the data is even, the median is computed in the following manner:

58 34 32 47 74 67 35 34 30 39 (N= 10)

**Step 1**: First the data is to be arranged in either ascending or descending order. We will arrange the data in ascending order and it will look like this:

30 32 34 34 35 39 47 58 67 74

**Step 2**: The following formula is used to compute median:

$M_d = (N/2)^{th}$ score + $[(N/2)^{th}$ score + 1$]/2$

The $(N/2)^{th}$ score is the $5^{th}$ score, that is 35.

The $(N/2)^{th}$ score + 1 is the $6^{th}$ score, that is 39.Thus 35 +39/ 2= 37

The median thus obtained is 37.

### 3.5.5 Computation of Median for Grouped Data

The formula used for computation of median for grouped data is as follows:

$M_d = L + \left[(N/2) - F/f_m\right] \times i$

Where,

L = The lower limit of the median class

N = Total of all the frequencies

F = Sum of frequencies before the median class

$f_m$ = frequency within the interval upon which the median falls

i = class interval.

Let us discuss the steps followed for computation of median with the help of the example given below:

| Class Intervals (Marks) | Frequencies (*f*) |
|---|---|
| 35- 39 | 5 |
| 30-34 | 7 |
| **25-29** | **5** |
| 20-24 | 6 |
| 15-19 | 4 |
| 10- 14 | 3 |
| | **N= 30** |

The steps in computing median for grouped data are as follows:

**Step 1:** The first step is to compute N/2, that is 30/2 so that we obtainone half of the scores in the data (15 in this case).

**Step 2:** As the scores are even in number (N= 30), the median should fall between 15th and 16th score. Whether we add the frequencies from above (5+7+5= 17) or from below (3+4+6+5= 18), the median will fall in the class interval 25-29. Further L that is the lower limit of the median class can also be mentioned. As the median class is 25-29, its lower limit will be 24.5.

**Step 3:** Compute F, that is sum of frequencies before the median class. In our example it would be 3 +4 +6 = 13

**Step 4:** $f_m$ is computed. It is the frequency within the interval upon which the median falls. In the present example the median class interval is 25-29 and the frequency for this class interval is 5. So $f_m$ is 5.

**Step 5:** The values can now be put in the formula to obtain the median

$$M_d=L+[(N/2)-F/f_m] \times i$$

$$= 24.5 +[(30/2)- 13/ 5] \times 5$$

$$= 24.5 + [15- 13/5] \times 5$$

$$= 24.5 + [2/ 5] \times 5$$

$$= 24.5 + 10/ 5$$

$$= 24.5 +2$$

$$= 26. 5$$

Thus, the median obtained is 26. 5. And it falls in the median class interval 25-29.

### 3.5.6 Computation of Mode for Ungrouped Data

Let us now learn how to compute mode for an ungrouped data with the help of the following example:

58 34 32 47 74 67 35 34 30 39

The mode can be calculated in simple manner by just counting the scores that appears maximum number of times in the data. In our example, the score occurring maximum number of times is 34, that occurs twice. Thus the mode is 34.

### 3.5.7 Computation of Mode for Grouped Data

There are two methods by which mode for grouped data can be computed:

#### 3.5.7.1 First Method

The first method is by using the following formula

**Mode=3Mdn−2M**

Where,

Mdn = Median

M= Mean

Let us now compute mode with the help of the following example:

| **Class Intervals (Marks)** | **Frequencies (*f*)** | **Midpoint (X)** | ***f*X** |
|---|---|---|---|
| 50- 59 | 5 | 54.5 | 272.5 |
| 40- 49 | 7 | 44.5 | 311.5 |
| 30- 39 | 8 | 34.5 | 276 |
| 20- 29 | 10 | 24.5 | 245 |
| 10- 19 | 15 | 14.5 | 217.5 |
| 0- 9 | 5 | 4.5 | 54.522.5 |
| | **N= 50** | | **ΣfX = 1345** |

The formula M = Σ*f*X/ N is used, N is equal to 50.

**Step 1**: Compute mean

**M = Σ*f*X/ N**

M = 1370/ 50= 26.9

**Step 2**: Compute median

$$\mathbf{M_d = L + [(N/2) - F/f_m] \times i}$$

$= 19.5 + [(50/2) - 20/10] \times 10$

$= 19.5 + [25 - 20/10] \times 10$

$= 19.5 + [5/10] \times 10$

$= 19.5 + 5$

$= 24.5$

**Step 3**: Let us now use these values in our formula and compute mode

$\mathbf{M_o = 3Mdn - 2M}$

$M_o = 3 \times 24.5 - 2 \times 26.9$

$= 73.5 - 53.8$

$= 19.7$

Thus the mode computed is 19.7

Also we can make one observation here that the mean obtained for our example is 26.9 the median is 24.5 and the mode is 19.7. All the three values are not close to each other indicating that the distribution of the data may not be normal as the values do not fall in the central area of the distribution. If the values of mean, median and mode were similar, then we could have said that the data is normally distributed.

### 3.5.7.2 Second Method

In the second method of computing mode for grouped data the following formula is used:

$$\mathbf{M_o = L + [d_1 / d_1 + d_2] \times i}$$

Where,

L= Lower limit of the class interval in which the mode may lie, called as modal class

i = Class interval

$d_1$ = difference between frequencies of modal class and class interval below it.

$d_2$ = difference between frequencies of modal class and class interval above it,

Let us discuss the steps followed for computation of mode with the help of the example given below:

| Class Intervals (Marks) | Frequencies (*f*) |
|---|---|
| 35- 39 | 5 |
| **30-34** | **7** |
| 25-29 | 5 |
| 20-24 | 6 |
| 15-19 | 4 |
| 10- 14 | 3 |
| | **N= 30** |

**Step 1**: The mode is most likely to fall in the the class intervals 30-34 as that has the highest frequencies (7). Thus this is our modal class and the lower limit of the same (L) will be 29.5.

**Step 2**: The class interval (i) for this example is 5.

**Step 3**: Compute $d_1$, that is, difference between frequencies of modal class and class interval below it and $d_2$, that is, difference between frequencies of modal class and class interval below it.

$d_1 = f_m - f_{m-1}$

$d_2 = f_m - f_{m+1}$

Where,

$f_m$ = the frequency of the modal class (7 in case of our example).

$f_{m-1}$ = the frequency of the class interval below the modal class (5 in case of our example).

$f_{m+1}$ = the frequency of the class interval above the modal class (5 in case of our example).

Thus, $d_1$= 7-5 = 2 and $d_2$= 7-5 = 2 in case of our example.

**Step 4**: Now let us compute mode with the help of the formula

$\mathbf{M_o = L + [d_1 / d_1 + d_2] \times i}$

$M_o$= 29.5 +[2/2+2]x 5

= 29.5 +2/4 x 5

= 29.5 +10 /4

= 29.5 + 2.5

= 32

Thus, the mode obtained is 32.

**Check Your Progress IV**

1) Compute mean, median and mode for the following data:

23, 34, 43, 65, 67, 67, 78, 65, 43, 34, 45, 33, 23, 67, 60 (N= 15)

2) Compute mean for the following data:

| Class Intervals (Marks) | Frequencies (*f*) |
|---|---|
| 50- 59 | 4 |
| 40- 49 | 5 |
| 30- 39 | 6 |
| 20- 29 | 5 |
| 10- 19 | 5 |
| 1- 9 | 5 |
| | **N= 30** |



## 3.6 LET US SUM UP

In the present unit, we discussed the concept of central tendency. The measures of central tendency was explained as summary figures that help in describing a central location for a certain group of scores. It was further explained as providing information about the characteristics of the data by identifying the value at or near the central location of the data. The functions of measures of tendency besides the characteristics of good measures of central tendency were also discussed. Further, the unit focused on the three measures of central tendency, namely, mean, median and mode. Mean is a total of all the scores in data divided by the total number of scores. It is one of the most frequently used measure of central tendency and is often referred to as an average. It can also be termed as one of the most sensitive measure of central tendency as all the scores in a data are taken in to consideration when it is computed. Median is the middle score in an ordered distribution. Median is a point in any distribution below and above which lie half of the scores. Mode is the score in a distribution that occurs most frequently. Certain distributions are bimodal, where there are two modes.When there are three modes, the term used is trimodal and when there are four or more modes, we use the term multimodal. Though, if the scores in a distribution greatly vary, then it is possible that there is no mode. The properties, advantages and limitations of mean, median and mode were also discussed in detail. Further, the computation of each of these measures of central tendency was also discussed for both ungrouped and grouped data with stepwise explanation.

## 3.7 REFERENCES

Bordens, K. S. and Abbott, B. B. (2011). *Research Design and Methods: A Process Approach*. New Dekhi:McGraw Hill Education(India) Private Limited.

King, Bruce. M; Minium, Edward. W. (2008). *Statistical Reasoning in the Behavioural Sciences*. Delhi: John Wiley and Sons, Ltd.

Mangal, S. K. (2002). *Statistics in Psychology and Education*. new Delhi: Phi Learning Private Limited.

Minium, E. W., King, B. M., & Bear, G. (2001). *Statistical Reasoning in Psychology and Education*. Singapore: John-Wiley.

Mohanty, B and Misra, S. (2016). *Statistics for Behavioural and Social Sciences*. Delhi: Sage.

Tate, M. W.(1955). *Statistics in Education*. New York: Macmillion Co.

Veeraraghavan, V and Shetgovekar, S. (2016). *Textbook of Parametric and Nonparametric Statistics*. Delhi: Sage.

## 3.8 KEY WORDS

**Measures of Central Tendency:** Measures of central tendency can be explained as a summary figure that helps in describing a central location for a certain group of scores.

**Mean:** Mean is a total of all the scores in data divided by the total number of scores.

**Median:** Median is a point in any distribution below and above which lie half of the scores.

**Mode:** Mode is the score in a distribution that occurs most frequently.

## 3.9 ANSWERS TO CHECK YOUR PROGRESS

**Check Your Progress I**

1) Define Measures of Central Tendency

Measures of central tendency can be defined as a summary figure that helps in describing a central location for a certain group of scores. It is a value that determines the general magnitude of a distribution.

1) List the functions of measures of central tendency.

   a) They provide a summary figure with the help of which the central location of the whole data can be explained.

   b) The large amount of data can be easily reduced to a single figure.

   c) When mean is computed for a certain sample, it will help us gain idea about the population mean.

   d) The results obtained from computing measures of central tendency will help a researcher make certain decisions.

   e) Comparison can be carried out with the help of the single figures computed with the help of measures of central tendency.

**Check Your Progress II**

1) Describe mean, median and mode with suitable examples.

| Measure | Description | Example |
|---|---|---|
| **Mean** | Mean is a total of all the scores in data divided by the total number of scores. It is one of the most often used measures of central tendency and is often referred to as average. It can also be termed as one of the most sensitive measures of central tendency as all the scores in a data are taken in to consideration when it is computed. | Scores on Job Satisfaction obtained by 5 employees<br>23, 34, 54, 34, 22 (N= 5)<br>Thus Mean would be<br>23 +34 +54 + 34 +22= 167<br>Thus 167/5= 33.4 |
| **Median** | Median is the middle score in an ordered distribution. Median is a point in any distribution below and above which lie half of the scores. | In above example, the data is arranged in ascending order<br>22, 23, 34, 34, 54<br>Median thus is 34 |
| **Mode** | Mode is the score in a distribution that occurs most frequently. Certain distributions can be bimodal as well, where there are two modes.When there are three modes, the term used is trimodal and when there are four or more modes, we use the term multimodal. Though if the scores in a distributions greatly vary then it is possible that there is no mode. | In above example,<br>23, <u>34,</u> 54, <u>34</u>, 22<br><br>Mode is 34 that occurs twice |

2) How to choose which measure of central tendency to use?

Choice of measure of central tendency will depend on the scales of measurement and also whether the data is normally distributed or not.

**Check Your Progress III**

1) List the properties of mean

   a) Mean is sensitive to the actual position of each and every score in a distribution and if another score is included in the distribution, then the mean or average of that distribution will change.

   b) Mean denotes a balance point of any distribution and the total of positive deviations from the mean is equal to the negative deviations from the mean.

   c) Mean is especially effective when we want the measure of central tendency to needs to reflect the sum of the scores.

2) List the advantages of median.

   a) The definition of median is rigid which is a quality of a good measure of central tendency.

   b) It is easy to understand and calculate.

   c) It is not affected by outliers or extreme scores in data.

   d) Unless the median falls in an open ended class, it can be computed for grouped data with open ended classes.

   e)I n certain cases it is possible to identify median through inspection as well as graphically.

3) List the limitations of mode.

   a) It is sometimes possible that the scores in the data vary from each other and in such cases the data may have no mode.

   b) Mode cannot be rigidly defined.

   c) In case of bimodal, trimodal or multimodal distribution, interpretation and comparison becomes difficult.

   d) Mode is not based on the whole distribution.

   e) It may not be possible to compute further mathematical procedures based on mode.

   f) Sampling fluctuations can have an impact on mode.

**Check Your Progress IV**

1) Compute mean, median and mode for the following data:

   23, 34, 43, 65, 67, 67, 78, 65, 43, 34, 45, 33, 23, 67, 60 (N= 15)

   Mean = 49.8, Median = 45, Mode: 67

2) Compute mean for the following data:

| Class Intervals (Marks) | Frequencies (*f*) |
|---|---|
| 50- 59 | 4 |
| 40- 49 | 5 |
| 30- 39 | 6 |
| 20- 29 | 5 |
| 10- 19 | 5 |
| 1- 9 | 5 |
| | **N= 30** |

**Mean = 28.83**

## 3.10 UNIT END QUESTIONS

1) Discuss the concept of measures of central tendency with a focus on characteristics of a good measure of central tendency.

2) Explain the properties of mean, median and mode.

3) Discuss the limitations of mean, median and mode.

4) Compute mean, median and mode for the following data:

   44, 32, 34, 34, 45,54, 56, 54, 55, 58, 45, 56, 54, 55, 56, 67, 79, 77, 88, 66, 89, 65, 43, 45, 54

5) Compute mean, median and mode for the following data:

| Class Intervals (Marks) | Frequencies (*f*) |
|---|---|
| 50- 59 | 12 |
| 40- 49 | 10 |
| 30- 39 | 9 |
| 20- 29 | 11 |
| 10- 19 | 8 |
| 1- 9 | 10 |
| | **N= 60** |

# UNIT 4 INTRODUCTION TO MEASURES OF VARIABILITY*

**Structure**



## 4.0 OBJECTIVES

After reading this unit, you will be able to:

- explain the concept of variability in data;
- describe the main properties limitation and uses of the range, quartile deviation, average deviation and standard deviation; and
- explain variance and coefficient of variance.

## 4.1 INTRODUCTION

Look at the two data given below:

Data A: 8, 2, 6, 4, 8, 2, 10, 5, 5, 10 (N= 10, Total = 60, Mean = 6)

Data B: 7, 7, 7, 6, 7, 5, 5, 6, 5, 5 (N = 10, Total = 60, Mean = 6)


* Dr. Usha Kulshreshtha, Faculty, Psychology, University of Rajasthan, Jaipur.

A single glance at the data A and B given above tell us that data A is more homogeneous when compared to the data B that seems to display more variability. Though to further understand the variance in the data, various measures of variability need to be computed.

In the previous unit, we discussed the measures of central tendency, viz, mean, median and mode. These measures give us an average of a set of observations or data. However, the average cannot be a true representation of data because of variations in the distribution. As can be seen in the above example, the mean is same for data A and data B, but the data vary in terms of their deviation from the mean. Thus, it is very important to consider the variations in the data or set of observations. In this unit, we will be explaining the concept of variability (also known as dispersion) in data. Dispersion actually refers to the variations that exist within and amongst the scores obtained by a group. In average there is a convergence of scores towards a mid-point in a normal distribution. In dispersion, we try and see how each score in the group varies from the mean or the average score. The larger the dispersion, less is the homogeneity of the group concerned and if the dispersion is less it means that the group is homogeneous. Dispersion is an important statistic which helps us to know how far the sample population varies from the universe population. It tells us about the standard error of the mean.

In the present unit, we will discuss the meaning and significance of variability. The main properties and limitation of the range, quartile deviation, average deviation and standard deviation that are the measures of variability will also be discussed. Further, the concept of variance and coefficient of variance will also be highlighted.

## 4.2 CONCEPT OF VARIABILITY IN DATA

Variability in statistics means deviation of scores in a group or series, from their mean scores. It actually refers to the spread of scores in the group in relation to the mean. It is also known as dispersion. For instance, in a group of 10 participants who have scored differently on a mathematics test, each individual varies from the other in terms of the marks that he/she has scored. These variations can be measured with the help of measure of variability, that measure the dispersion of different values for the average value or average score. Variability or dispersion also means the scatter of the values in a group. High variability in the distribution means that scores are widely spread and are not homogeneous. Low variability means that the scores are similar and homogeneous and are concentrated in the middle.

According to Minium, King and Bear (2001), measures of variability express quantitatively the extent to which the score in a distribution scatter around or cluster together. They describe the spread of an entire set of scores, they do not specify how far a particular score diverges from the centre of the group. These measures of variability do not provide information about the shape of a distribution or the level of performance of a group.

Measures of variability fall under descriptive statistics that describe how similar a set of scores are to each other. The greater the similarity of the scores to each other, lower would be the measure of variability or dispersion. The less the similarity of the scores are to each other, higher will be the measure of variability or dispersion. In general, the more the spread of a distribution, larger will be the measure of dispersion. To state it succinctly, the variation between the data values in a sample is called dispersion. The most commonly used measures of dispersion are the range, and standard deviation.

In the previous unit, measures of central tendency were discussed. While measures of central tendencies are indeed very valuable, their usefulness is rather limited. Although through these measures we can compare the two or more groups, a measure of central tendency is not sufficient for the comparison of two or more groups. They do not show how the individual scores are spread out. Let us take another example, similar to the one that we discussed under the section on introduction. A math teacher is interested to know the performance of two groups (A and B) of his /her students. He/she gives them a test of 40 points. The marks obtained by the students of groups A and B in the test are as follows:

Marks of Group A: 5,4,38,38,20,36,17,19,18,5 (N = 10, Total = 200, Mean = 20)

Marks of Group B: 22,18,19,21,20,23,17,20,18,22 (N = 10, Total = 200, Mean = 20)

The mean scores of both the groups is 20, as far as mean goes there is no difference in the performance of the two groups. But there is a difference in the performance of the two groups in terms of how each individual student varies in marks from that of the other. For instance, the test scores of group A are found to range from 5 to 38 and the test scores of group B range from 18 to 23.

It means that some of the students of group A are doing very well, some are doing very poorly and performance of some of the students is falling at the average level. On the other hand, the performance of all the students of the second group is falling within and near about the average (mean) that is 20. It is evident from this that the measures of central tendency provide us incomplete picture of a set of data. It gives insufficient base for the comparison of two or more sets of scores. Thus, in addition to a measure of central tendency, we need an index of how the scores are scattered around the center of the distribution. In other words, we need a measure of dispersion or variability. A measure of central tendency is a summary of scores, and a measure of dispersion is summary of the spread of scores. Information about variability is often as important as that about the central tendency.

The term variability or dispersion is also known as the *average of the second degree*, because here we consider the arithmetic mean of the deviations from the mean of the values of the individual items. To describe a distribution adequately, therefore, we usually must provide a measure of central tendency and a measure of variability. Measures of variability are important in statistical inference. With the help of measures of dispersion, we can know about fluctuation in random sampling. How much fluctuation will occur in random sampling? This question in fundamental to every problem in statistical inference, it is a question about variability.

The measures of variability are important for the following purposes:

- Measures of variability are used to test the extent to which an average represents the characteristics of a data. If the variation is small then it indicates high uniformity of values in the distribution and the average represents the characteristics of the data. On the other hand, if variation is large then it indicates lower degree of uniformity and unreliable average.

- Measures of variability help in identifying the nature and cause of variation. Such information can be useful to control the variation.
- Measures of variability help in the comparison of the spread in two or more sets of data with respect to their uniformity or consistency.
- Measures of variability facilitate the use of other statistical techniques such as correlation, regression analysis, and so on.

### 4.2.1 Functions of Variability

The major functions of dispersion or variability are as follows:

- It is used for calculating other statistics such as analysis of variance, degree of correlation, regression etc.
- It is also used for comparing the variability in the data obtained as in the case of Socio-Economic Status, income, education etc.
- To find out if the average or the mean/median/mode worked out is reliable. If the variation is small then we could state that the average calculated is reliable, but if variation is too large, then the average may be erroneous.
- Dispersion gives us an idea if the variability is adversely affecting the data and thus helps in controlling the variability.

### 4.2.2 Absolute Dispersion and Relative Dispersion

Measures of dispersion give an estimate and express quantitatively the deviation of individual scores in a given sample from the mean and median. Thus, the numerical measures of variability spread or scatter around a central value.

In measuring dispersion, it is imperative to know the amount of variation (absolute measure) and the degree of variation (relative measure). In the former case, we consider the range, mean deviation, standard deviation etc. In the latter case, we consider the coefficient of range, the coefficient of mean deviation, the coefficient of variation etc. Thus, there are two broad classes of the measures of dispersion or variability. They are absolute measure of dispersion and relative measure of dispersion.

Absolute dispersion usually refers to the standard deviation, a measure of variation from the mean. The units of standard deviation are the same as for the data. In other words, absolute measure is expressed in terms of the original units of a distribution. Therefore, absolute dispersion is not suitable for comparing the variability of two distributions since the two variables are expressed and measured in two different units. For instance, the variability in body height (cm) and body weight (kg) cannot be compared because the absolute measure (standard deviation) is expressed in cm and kg. The absolute measure is also not appropriate for two sets of scores expressed in the same units with wide divergence in means (central value). Nevertheless, absolute measures are widely used, except in the exceptional cases like above. The absolute measures include range, mean deviation, standard deviation, and variance.

Relative dispersion, sometimes called the coefficient of variation, is the result of dividing the standard deviation by the mean and it may be presented as a quotient or as a percentage. Thus, relative measures are computed from the absolute measures of dispersion and its corresponding central values. A low value of relative dispersion usually implies that the standard deviation is small in comparison to the magnitude of the mean. To give an example, if standard deviation for mean of 30 marks is 6.0, then the coefficient of variation will be

$$6.0/ 30 = 0.2 \text{(about 20\%)}$$

If the mean is 60 marks and the standard deviation remains the same as 6.0, the coefficient of variation will be

$$6.0 / 60 = 0.1 \ (10\%).$$

However, with measurements on either side of zero and a mean being close to zero the relative dispersion could be greater than 1. At the same time, we must remember that the two distributions in quite a few cases can have the same variability. Sometimes the distributions may be skewed and not normal with mean, mode and median at different points in the continuum. These distributions are called skewed distributions (Skewness will be discussed in detail in the unit 8). It is also possible to have two distributions that have equal variability but unequal means or different shapes. Thus, the relative measure is derived from a ratio of an absolute measure like standard deviation and mean (measure of central value) and is expressed in percentage of the mean. So, the relative measure is suitable for comparing the variabilities of two sets of scores given in different units. They are also preferred in comparing two sets of scores given in the same unit, when the mean widely diverges. The relative measures include the coefficient of variation, the coefficient of quartile deviation, and the coefficient of mean deviation.

**Check Your Progress I**

1) State any one function of variability.

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

2) List the two broad classes of the measures of dispersion.

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

.....................................................................................................................

## 4.3 TYPES OF MEASURES OF DISPERSION OR VARIABILITY

The measures of variability most commonly used in psychological statistics are as follow:

1) Range

2) Quartile Deviation

3) Average Deviation or Mean Deviation

4) Standard Deviation

5) Variance

Range and quartile deviation measure dispersion by computing the spread within which the values fall, while as average deviation and standard deviation compute the extent to which the values differ from the average. We will introduce each and discuss their properties in detail.

## 4.3.1 The Range (R)

Range can be defined as the difference between the highest and lowest score in the distribution. This is calculated by subtracting the lowest score from the highest score in the distribution. The equation is as follows:

Range = Highest Score – Lowest Score(R=H-L)

The range is a rough measure of dispersion because it tells about the spread of the extreme scores and not the spread of any of the scores in between. For instance, the range for the distribution 4,10,12,20, 25, 50 will be 50 - 4 = 46.

### 4.3.1.1 Merits and Limitations of the Range

Some of the merits of range as a measure of variability are explained in this section.

1) It is easiest to compute when compared with other measures of variability and its meaning is direct.

2) The range is ideal for preliminary work or in other circumstances where precision is not an important requirement (Minium et. al., 2001).

3) It is quite useful in case where the purpose is only to find out the extent of extreme variation, such as temperature, rainfall etc.

4) Range is effectively used in the application of tests of significance with small samples.

The following are the limitations of the range as a measure of variability:

1) The calculation of range is based only on two extreme values in the data set and does not consider other values of the data set. Sometimes, the extreme values of the two different data sets may be same or similar, but the two data sets may be differ in dispersion.

2) Its value is sensitive to change in sampling. The range varies more with sampling fluctuation. That is different sample of the same size from the same population may have different range.

3) Its value is influenced by large samples. In many types of distribution, including normal distribution, the range is dependent on sample size. The sampling variance increases rapidly with increase in sample size.

4) Range cannot be used for open-ended class intervals since the highest and the lowest scores of the distribution are not available and thus the range cannot be computed.

5) Further mathematical calculations are not possible for range.

6) Range indicates two extreme scores, thus the magnitude or frequency of intermediate scores is missing.

7) It does not indicate the form of distribution, like skewness, kurtosis, or modal distribution of scores.

8) A single extreme score may also increase the range disproportionately.

#### 4.3.1.2 Uses of the Range

Range is applied in diverse areas discussed as follows:

- Range is used in areas where there are small fluctuations, such as stock market, rate of exchange, etc.
- Range may be used in day-to-day activities like, daily sales in a grocery store, monthly wages in a factory, etc.
- Range is used in weather forecasts, like variation in temperature in a day.
- When the researcher is only interested in the extreme scores or total spread of the scores, range is the most useful measure of variability.
- Range can also be used when the data are too scant or too scattered to justify the use of most appropriate measure of variability.

### 4.3.2 The Quartile Deviation (QD)

Since a large number of values in the data lie in the middle of the frequencies distribution and range depends on the extreme (outliers) of a distribution, we need another measure of variability. The Quartile deviation, is a measure that depends on the relatively stable central portion of a distribution. According to Garret (1966), the Quartile deviation is half the scale distance between $75^{th}$ and $25^{th}$ per cent in a frequency distribution. The entire data is divided into four equal parts and each part contains 25% of the values. According to Guilford (1963) the Semi-Interquartile range is the one half the range of the middle 50 percent of the cases.

On the basis of above definitions, it can be said that quartile deviation is half the distance between $Q_1$ and $Q_3$.

**Inter Quartile Range (IQR):** The range computed for the middle 50% of the distribution is the interquartile range. The upper quartile ($Q_3$) and lower quartile ($Q_1$) is used to compute IQR. This is $Q_3 – Q_1$. IQR is not affected by extreme values.

**Semi-Interquartile Range (SIQR) or Quartile Deviation (QD):** Half of the IQR is called as semi inter quartile range. SIQR is also called as quartile deviation or QD. Thus, QD is computed as;

$$QD = Q_3 - Q_1/2$$

Thus, quartile deviation is obtained by dividing IQR by 2. Quartile deviation is an absolute measure of dispersion and is expressed in the same unit as the scores.

Quartile deviation is closely related to the median because median is responsive to the number of scores lying below it rather than to their exact positions and $Q_1$ and $Q_3$ are defined in a same manner. The median and quartile deviation have common properties. Both median and quartile deviation are not affected by extreme values. In a symmetrical distribution, the two quartiles $Q_1$ and $Q_3$ are at equal distance from the median or $Q_1$ =$Q_3$- Median. Thus, like median, quartile deviation covers exactly 50 per cent of observed values in the data. In normal distribution, quartile deviation is called the Probable Error or PE. If the distribution is open-class, then quartile deviation is the only measure of variability that is reasonable to compute.

In an asymmetric or skewed distribution, $Q_1$ and $Q_3$ are not equidistant from $Q_2$ or median. In such a distribution, the median of the IQR moves towards the skewed tail. The degree and direction of skewness can be assessed from quartile deviation and the relative distance between $Q_1$, $Q_2$ and $Q_3$.

Kurtosis is proportional to quartile deviation. Smaller the quartile deviation, greater is the concentration of scores in the middle of the distribution, thus making the distribution with high peak and narrow body. The scores that are widely dispersed indicate a large quartile deviation and thus, long IQR. This distribution has a low peak and broad body.

#### 4.3.2.1 Merits and Limitations of Quartile Deviation

From the explanation in the above section, it becomes clear that quartile deviation is easy to understand and compute.

1) Quartile deviation is a better measure of dispersion than range because it takes into account 50 per cent of the data, unlike the range which is based on two values of the data, that is highest value and the lowest value.

2) Secondly, quartile deviation is not affected by extreme scores since it does not consider 25 per cent data from the beginning and 25 per cent from the end of the data.

3) Lastly, quartile deviation is the only measure of dispersion which can be computed from the frequency distribution with open-end class.

Despite the major merits of quartile deviation, there are limitations to it as well.

1) The value of quartile deviation is based on the middle 50 percent values, it is not based on all the observations. Thus, it is not regarded as a stable measure of variability

2) The value of quartile deviation is affected by sampling fluctuation.

3) The value of quartile deviation is not affected by the distribution of the individual values within the intervals of middle 50 percent observed values.

#### 4.3.2.2 Uses of Quartile Deviation

1) The distribution contains few and very extreme scores.

2) When the median is the measure of central tendency.

3) When our primary interest is to determine the concentration around the median.

### 4.3.3 The Average Deviation (AD) or Mean Deviation (MD)

The two measures of variation, range and quartile deviation, which we discussed in the earlier subsections, do not show how values of the data are scattered about a central value. R and QD attempt to compute spread of values and not compute how far the values are from their average. To measure the variation, as a degree to which values within a data deviate from their mean, we use average deviation.

Before discussing average deviation, first we should know about the meaning of deviation. Deviation score express the location of the scores by indicating how many score points it lies above or below the mean of the distribution. Deviation score may be defined as (X-Mean), that is, when we subtract the means from each of the raw scores the resulting deviation scores states the position of the scores, relative to the mean.

According to Garrett (1971, as cited in Mangal 2002, page 70) "The average deviation is the mean of the deviation of all of the separate scores is a series taken from their mean". According to Guilford (1963) average deviation can be described as an average or mean of all the deviations when the algebraic signs are not taken in to the account.

Average is a central value and thus, some deviations will be positive (+) and some may be negative (-). Mean deviation ignores the signs of the deviations, and it considers all the deviations to be positive. This is so because the algebraic sum of all the deviations from the mean equals to zero. MD or AD is arithmetic mean of the difference of the values from the average. The average is either the arithmetic mean or the median. It is a measure of variability that takes into account the variations of all the scores in the data. It is an absolute measure of dispersion and is expressed in the same unit as the raw scores.

The calculation of average deviation is easy therefore it is a popular measure. When we calculate average deviation, equal weight is given to each observed value and thus it indicates how far each observation lies from the mean. AD or MD can be obtained from any of the measures of central tendency, that is mean, median, or mode. Mode is ill defined, hence, AD or MD is computed about the mean or median. AD or MD calculated about the median will be less than the AD or MD about the mean or mode. For a symmetrical distribution, MD about mean and MD about median covers 57.5 per cent of the observations of the data. Thus, a small value of MD will indicate less variability. AD is thus somewhat larger (57.5 per cent of the cases) than QD (50 per cent of the cases).

#### 4.3.3.1 Merits and Limitations of the Average Deviation

The main merits of AD are as follows:

1) AD or MD is easy to understand and compute.

2) It is based on all observations, unlike R or QD.

3) It is an accurate measure of variability since it averages the absolute deviations.

4) It is less affected by extreme observations.

5) It is based on average thus, it is a better measure to compare about the formations of different distributions.

The main limitations of average deviation are as follows:

1) While calculating average deviation we ignore the plus minus sign and consider all values as plus. Because of this mathematical property, it is not used in inferential statistics.

2) AD cannot be computed for open-end classes.

3) It tends to increase with the size of the sample.

#### 4.3.3.2 Use of Average Deviation

Despite the limitations, AD or MD is used by economists and business statisticians. It is also used in computing the distribution of personal wealth in a community or a nation. According to National Bureau of Economic Research, MD is the most practical measure of dispersion to be used for this purpose (Mohanty and Misra, 2016, pg. 133).

1) When it is desired to weight all deviation from the mean according to their size.

2) When the standard deviation in unduly influenced by the presence of extreme scores.

3) Distribution of the score is not near to normal.

### 4.3.4 The Standard Deviation (SD)

The term standard deviation was first used in writing by Karl Pearson in 1894. The standard deviation of population is denoted by ‘σ’ (Greek letter sigma) and that for a sample is ‘s’**.** A useful property of SD is that unlike variance it is expressed in the same unit as the data. This is most widely used method of variability. The standard deviation indicates the average of distance of all the scores around the mean. It is the positive square root of the mean of squared deviations of all the scores from the mean. It is the positive square root of variance. It is also called as ‘root mean square deviation’. Mangal (2002, page 71) defined standard deviation as “as the square root of the average of the squares of the deviations of each score from the mean”. SD is an absolute measure of dispersion and it is the most stable and reliable measure of variability.

Standard deviation shows how much variation there is, from the mean. SD is calculated from the mean only. If standard deviation is low it means that the data is close to the mean. A high standard deviation indicates that the data is spread out over a large range of values. Standard deviation may serve as a measure of uncertainty. If you want to test the theory or in other word, want to decide whether measurements agree with a theoretical prediction, the standard deviation provides the information. If the difference between mean and standard deviation is very large then the theory being tested probably needs to be revised. The mean with smaller standard deviation is more reliable than mean with large standard deviation. A smaller SD shows the homogeneity of the data. The value of standard deviation is based on every observation in a set of data. It is the only measure of dispersion capable of algebraic treatment therefore, SD is used in further statistical analysis.

#### 4.3.4.1 Merits and Limitations of the Standard Deviation

The main merits of using standard deviation are as follows:

1) It is widely used because it is the best measure of variation by virtue of its mathematical characteristics.

2) It is based on all the observations of the data.

3) It gives an accurate estimate of population parameter when compared with other measures of variation.

4) SD is least affected by sample fluctuations

5) It is also possible to calculate combined SD, that is not possible with other measures.

6) Further statistics can be applied on the basis of SD like, correlation, regression, tests of significance, etc.

7) Coefficient of variation is based on mean and SD. It is the most appropriate method to compare variability of two or more distributions.

The limitations of SD are as follows:

1) While calculating standard deviation more weight is given to extreme values and less to those, near the means. When we calculate SD, we take deviation from mean (X-M) and square these obtained deviations. Therefore, large deviations, when squared are proportionally more than small deviations. For example, the deviations 2 and 10 are in the ratio of 1:5 but their square 4 and l00 are in the ratio 1:25.

2) It is difficult to compute as compared to other measures of dispersion.

#### 4.3.4.2 Uses of Standard Deviation

The uses of standard deviation are as follows:

1) SD is used when one requires a more reliable and accurate measure of variability but it is recommended when the distribution is normal or near to normal.

2) It is used when further statistics like, correlation, regression, tests of significance, etc. have to be computed.

### 4.3.5 Variance

The term variance was used to describe the square of the standard deviation by R.A. Fisher in 1913. The concept of variance is of great importance in advanced work where it is possible to split the total into several parts, each attributable to one of the factors causing variations in their original series. Variance is a measure of the dispersion of a set of data points around their mean value. It is a mathematical expectation of the average squared deviations from the mean. The variance ($s^2$) or mean square (MS) is the arithmetic mean of the squared deviations of individual scores from their means. In other words, it is the mean of the squared deviation of scores.Variance is expressed as V = $SD^2$.

The variance and the closely related standard deviation are measures that indicate how the scores are spread out in a distribution. In other words, they are measures of variability. The variance is computed as the average squared deviation of each number from its mean.

Calculating the variance is an important part of many statistical applications and analysis. It is a good absolute measure of variability and is useful in computation of Analysis of Variance (ANOVA) to find out the significance of differences between sample means.

#### 4.3.5.1 Merits and Demerits of Variance

The main merits of variance are listed as follows:

1) It is rigidly defined and based on all observations.
2) It is amenable to further algebraic treatment.
3) It is not affected by sampling fluctuations.
4) It is less erratic.

The main demerits of variance are listed as follows:

1) It is difficult to understand and calculate.
2) It gives greater weight to extreme values.

#### 4.3.5.2 Co-efficient of Variation (CV)

The relative measure corresponding to SD is the coefficient of variation. It is a relative measure of dispersion developed by Karl Pearson. When we want to compare the variations (dispersion) of two different series, relative measures of standard deviation must be calculated. This is known as co-efficient of variation or the co-efficient of SD. It is defined as the SD expressed as a percentage of the mean.The coefficient of variation represents the ratio of the standard deviation to the mean, and it is a useful statistic for comparing the degree of variation from one data series to another, even if the means are drastically different from each other. Thus, it is more suitable than SD or variance. It is given as a percentage and is used to compare the consistency or variability of two or more data series.

The formula for computing coefficient of variation is as follows:

$$\mathbf{V = 100 \times \sigma / M}$$

Where,

V = Variance

σ = Standard deviation

M = Mean

To understand the computation with the help of an example, If the standard deviation of marks obtained by 10 students in a class test in English is 10 and Mean is 79, then,

$V = 100 \times 10 / 79$

$= 1000 / 79$

$= 12.65$

**Check Your Progress II**

1) What is range?

..........................................................................................................

..........................................................................................................

..........................................................................................................

..........................................................................................................

..........................................................................................................

..........................................................................................................

2) List the merits of quartile deviation.

..........................................................................................................

..........................................................................................................

..........................................................................................................

..........................................................................................................

..........................................................................................................

..........................................................................................................

3) What is variance?

..........................................................................................................

..........................................................................................................

..........................................................................................................

..........................................................................................................

..........................................................................................................

..........................................................................................................

..........................................................................................................

## 4.4 LET US SUM UP

To summarise, the measures of central tendency are not sufficient to describe data. Thus, to describe distribution adequately, we must provide a measure of variability or dispersion.The measures of variability are summary figures that express quantitatively, the extent to which, scores in a distribution scatter around or cluster together. The measures of variability are range, quartile deviation, average deviation, standard deviation and variance. Range is easy to calculate and useful for preliminary work. But this is based on extreme items only, and does not consider intermediate scores. Thus, it is not useful as a descriptive measure. Quartile deviation is related to the median in its properties. It takes into consideration the number of scores lying above or below the outer quartile point but not to their magnitude. This is useful with open ended distribution.The average deviation takes into account the exact position of each score in the distribution. The means deviation gives a more precise measure of the spread of scores but is mathematically inadequate. The average deviation is less affected by sampling fluctuation.The standard deviation is the most stable measure of variability. Standard deviation shows how much the score departs from the mean. It is expressed in original scores unit. Thus, it is most widely used measure of variability in descriptive statistics.The variance ($s^2$) or mean square (MS) is the arithmetic mean of the squared deviations of individual scores from their means. In other words, it the mean of the squared deviation of scores.The relative measure corresponding to SD is the coefficient of variation. It is a useful measure of relative variation.

## 4.5 REFERENCES

Garrett, H.E. (1981), *Statistics in Psychology and Education*, (Tenth edition), Bombay, Vakils Feffer and Simons Ltd.

McBride, Dawn M. (2018). *The Process of Statistical Analysis in Psychology.* Sage. USA

Minium, E.W., King, B.M. & Bear. G (2001). *Statistical Reasoning in Psychology and Education* (3rd edition), Singapore, John Wiley & Sons, Inc.

Mohanty, B. & Misra, Santa (2016). *Statistics for Behavioural and Social Sciences.* Sage. New Delhi.

## 4.6 KEY WORDS

**Average Deviation or Mean Deviation:** A measure of dispersion that gives the average difference (ignoring plus and minus sign) between each item and the mean.

**Dispersion:** The spread or variability is a set of data.

**Deviation:** The difference between raw score and mean.

**Quartile Deviation:** A measure of dispersion that can be obtained by dividing the difference between $Q_3$ and $Q_1$ by two.

**Range:** Difference between the largest and smallest value in a data.

**Standard deviation:** The square root of the variance in a series.

**Variance:** Variance is a measure of the dispersion of a set of data points around their mean value. It is a mathematical expectation of the average squared deviations from the mean.

## 4.7 ANSWERS TO CHECK YOUR PROGRESS

**Check Your Progress I**

1) State any one function of variability

   Variability is used for calculating other statistics such as analysis of variance, degree of correlation, regression etc.

2) List the two broad classes of the measures of dispersion.

   Absolute dispersion

   Relative dispersion.

**Check Your progress II**

1) What is range?

   Range can be defined as the difference between the highest and lowest score in the distribution.

2) List the merits of quartile deviation.

   The merits of quartile deviation are as follows:

   - Quartile deviation is a better measure of dispersion than range because it takes into account 50 percent of the data, unlike the range which is based on two values of the data, that is highest value and the lowest value.

   - Secondly, quartile deviation is not affected by extreme scores since it does not consider 25 percent data from the beginning and 25 percent from the end of the data.

   - Lastly, quartile deviation is the only measure of dispersion which can be computed from the frequency distribution with open-end class.

3) What is variance?

Variance is a measure of the dispersion of a set of data points around their mean value. It is a mathematical expectation of the average squared deviations from the mean. The variance ($s^2$) or mean square (MS) is the arithmetic mean of the squared deviations of individual scores from their means. In other words, it is the mean of the squared deviation of scores.

## 4.8 UNIT END QUESTIONS

1) Explain the concept and significance of variability.

2) Discuss the merits and limitation of range and quartile deviation

3) List the merits and limitations of standard deviation.

4) Elucidate average deviation or mean deviation.

5) Explain coefficient of variance with example.

# UNIT 5 COMPUTATION OF MEASURES OF VARIABILITY*

**Structure**



## 5.0 OBJECTIVES

After reading this unit, you will be able to:

- compute range;
- compute quartile deviation, for ungrouped and grouped data;
- compute average deviation, for ungrouped and grouped data; and
- compute standard deviation, for ungrouped and grouped data and with the help of short method.

## 5.1 INTRODUCTION

In the previous two units, we discussed about measures of central tendency and measures of variability. We discussed that average like mean, median and mode condense the series into a single figure. These measures of central tendency tell us something about the general level of magnitude of the distribution but they fail to show anything further about the distribution. It is not fully representative of a population unless we know the manner in which the individual items scatter around it. A further description of the series is necessary if we are to gauge how representative the average is. To cite an


* Dr. Usha Kulshreshtha, Faculty, Psychology, University of Rajasthan, Jaipur.

example, in a country the average income may be very high, yet there may be great disparity in its distribution among people. As a result, a majority of the people may be living below the poverty line. When we want to make comparison between two groups, it is seen that at times the value of the means is the same in both the groups but there is a large difference between individual participants in the groups. This difference amongst the participants within the same group is termed as variation, that is within the groups the participants vary a great deal even though they have the same means. Therefore to make accurate and meaningful comparisons between groups, we should use variability along with central tendency. In the last unit we have learned about the concept of variability, different measures of variability, their merits, limitations and uses. In this unit, we will learn how to compute the range, quartile deviation, average deviation and standard deviation.

## 5.2 COMPUTING DIFFERENT MEASURES OF DISPERSION (VARIABILITY)

As explained in the last unit that there are four measures of computing variability or dispersion within a set of scores:

1) Range (R)
2) Quartile Deviation (QD)
3) Average Deviation (AD) or Mean Deviation (MD)
4) Standard Deviation (SD)

Each of the above measures of variability give us the degree of variability or dispersion by the use of a single number and tells us how the individual scores are spread throughout the distribution. In the following sections, we will discuss the methods of computation of the above measures of dispersion.

### 5.2.1 Range

Range is the difference between the highest and the lowest score for a group of participants whose scores are given.

The formula for Range is as follows:

**R=H-L**

Where,

H=Highest scores in the distribution

L=Lowest score in the distribution

Let us understand the steps in computation of range with the help of an example,

For example, if there are 10 students who have obtained marks in history as mentioned below:

50, 45, 42, 46, 55, 54, 59, 60, 62, 64

**Step 1:** Arrange the scores in ascending order.

42, 45, 46, 50, 54, 55, 59, 60, 62, 64

**Step 2:** Identify the lowest and the highest score in the data

In the above data, the lowest score is 42 and the highest score is 64.

**Step 3:** Compute range with the help of the following formula:

**R=H-L**

64–42=22.

Thus, the range obtained is 22.

## 5.2.2 Quartile Deviation (QD)

The inter-quartile range is a measure of dispersion and is equal to the difference between the third and first quartiles. Half of the inter-quartile range is called semi inter-quartile range or Quartile Deviation. The formula for computation of QD is

**QD =** $Q_3 - Q_1 / 2$

Where,

$Q_1$ = first quartile of the data

$Q_3$= third quartile of the data

Quartile is an additional way of disaggregating data. Each quartile represents one-fourth of an entire population or the group. The quartile deviation has an attractive feature that the range "median+QD" contains approximately 50% of the data. The quartile deviation is also an absolute measure of dispersion. Its relative measure is called coefficient of quartile deviation or semi inter-quartile range. It is defined by the relation;

Coefficient of quartile deviation = $(Q_3-Q_1) / (Q_3+Q_1)$

The quartile deviation is one half the scale distance between the 75$^{th}$ and 25$^{th}$ percentile in a frequency distribution. The 25$^{th}$ percentile or $Q_1$ is the first quartile on the score scale, the point below which lies 25% of the scores.The 75$^{th}$ percentile or $Q_3$ is the third quartile on the score scale, the point below which lie 75% of the scores. To find quartile deviation, we must first compute $Q_3$ and $Q_1$.

There are grouped data and ungrouped data in all cases and thus to compute quartile deviation, we have to find out first if it is a grouped data ore ungrouped data. We will first see how the quartile deviation is calculated from ungrouped data.

### 5.2.2.1 Calculation of Quartile Deviation for Ungrouped Data

Let us understand steps in computation of quartile deviation for ungrouped data with the help of the following example:

The scores obtained by students in Psychology class test are :

24, 25, 23, 26, 29, 30, 27, 35, 34, 36, 28

**Step 1:** Arrange the data in ascending order.

23, 24, 25, 26, 27, 28, 29, 30, 34, 35, 36

**Step 2:** Compute $Q_1$

$Q_1 = (N+1)/4^{th}$ position

N = 11

$Q_1 = 11+1/4^{th}$ position = 3rd position = 25

**Step 3:** Compute $Q_3$

$Q_3 = 3(N+1)/4^{th}$ position

N = 11

$Q_3 = 3(11+1)/4 = 9^{th}$ position = 34

**Step 4:** Compute QD with help of the following formula

$$\mathbf{QD = Q_3 - Q_1 /2}$$

In the case of our data $Q_3$ is 34 and $Q_{1/2}$ is 25

QD =34–25/2

= 9/ 2

= 4.5

### 5.2.2.2 Calculation of Quartile Deviation from Grouped Data

From the grouped data, Quartile Deviation can be computed with the following formula:

$QD = Q_3 - Q_1/2$

Further,

$Q_1 = l + i[(N/4 - cum\,fi)]/fq$

$Q_3 = l + i[(3N/4 - cum\,fi)]/fq$

Where,

$l$ = the exact lower limit of the interval in which the quartile falls.

i = the length of the interval

cum *f*i = cumulative *f* up to interval which contains the quartile*f*q = the *f* on the

interval containing the quartile.

Let us understand steps in computation of quartile deviation for grouped data with the help of the following example:

| Class intervals | Frequencies ($f$) | Cumulative frequencies |
|---|---|---|
| 195-199 | 1 | 50 |
| 190-194 | 2 | 49 |
| 185-189 | 4 | 47 |
| 180-184 | 5 | 43 |
| 175-179 | 8 | 38 |
| 170-174 | 10 | 30 |
| 165-169 | 6 | 20 |
| 160-164 | 4 | 14 |
| 155-159 | 4 | 10(1+3+2+4) |
| 150-154 | 2 | 6 (1+3+2) |
| 145-149 | 3 | 4 (1+3) |
| 140-144 | 1 | 1 |

**Step 1:** $Q_1$ is calculated using the formula given below:

$Q_1 = l+i[(N/4–cumfi)]/fq$

To locate the $Q_1$, we take N/4. In the above example N/4=12.5

$l$ = 159.5 (50/4 = 12.5 th item from down below falls in 160-164)

$f$i = 10 cumulated scores upto interval containing $Q_1$

$f$q = 4, the f on the interval on which $Q_1$ falls

i = 5 (Class Interval)

Substituting in formula we have,

$Q_1$=159.5+5{(12.5-10)}/4=162.62

**Step 2:** To calculate the third quartile, that is $Q_3$.

$Q_3 = l+i\ [(3N/4–cum\ fi)]/fq$

To locate the $Q_3$ we take 3xN/4 of our scores. In the above example, 3N/4 is 3×50/4 = 37.5

3/4N = 37.5 (37.5th item that falls in 175-179

$l$ = 174.5 is the exact lower limit of interval which contains $Q_3$

Cum $f$i = 30, sum of scores upto interval which contains $Q_3$

i=5

$f$q=8

$Q_3$=174.5+5(37.5-30)/8 =179.19

**Step 3:** Finally, substituting in formula, we have the quartile deviation as given below in the formula;

QD = $Q_3 – Q_2$ / 2

Q = {(179.19)–(162.62)}/2

= 8.28

Thus, quartile deviation for the above data is 8.28.

### 5.2.3 Average Deviation (AD)

Average deviation as well can be computed for ungrouped and grouped data.

#### 5.2.3.1 Computation of Average Deviation for Ungrouped Data

In case of ungrouped data, the average deviation is calculated by the following formula,

**AD=$\sum|x|$/N**

Where,

$\sum|x|$ = Total of deviation from mean

N= Total number of observations

Let us understand the steps in computation of average deviation for ungrouped data with the help of example given below:

Following are the scores obtained by five students on a test:

| Students | Scores | Deviation from the Mean \| x \| |
|---|---|---|
| 1 | 6 | -4 |
| 2 | 8 | -2 |
| 3 | 10 | 0 |
| 4 | 12 | 2 |
| 5 | 14 | 4 |
| | **Total=50** | **Total=12(Ignore signs)** |
| | **Mean=50/5=10** | |

**Step 1:** Compute mean with the help of formula M=$\sum$X/N

M=50/5 =10

**Step 2:** Compute deviation from the mean as has been computed in the third column in above example.

It is seen from the table, that the deviations (x) are =0,-2,-4, +2, +4

**Step 3:** Compute total for these deviations without considering the + and – signs. the total is obtained as 12.

**Step 4**: Use the formula to compute average deviation.

AD = $\sum|x|$/N

= 12/5

=2.4.

Thus, the average deviation is obtained as 2.4.

### 5.2.3.2 Calculation of Average Deviation for Grouped Data

The average deviation for grouped data can be computed by the following formula,

**AD= ∑│fx│/N**

Where,

∑│fx│ =Add all the fx without considering the + and – sign

N= Number of observations

The above formula and calculation of AD can be illustrated by the example given below.

Let us understand the steps in computation of average deviation for grouped data with the help of example given below:

| **Class Interval** | **Frequency (*f*)** | **Mid Point (X)** | ***f*X** | **x=M-X** | ***f*x** |
|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) |
| 110-114 | 3 | 112 | 336 (112×3) | 20.33 (112-91.67=20.33) | 60.99 (20.33×3) |
| 105-109 | 4 | 107 | 428 | 15.33 | 61.32 |
| 100-104 | 6 | 102 | 612 | 10.33 | 61.98 |
| 95-99 | 8 | 97 | 776 | 5.33 | 42.64 |
| 90-94 | 15 | 92 | 1380 | .33 | 4.95 |
| 85-89 | 10 | 87 | 870 | –4.67 | –46.67 |
| 80-84 | 7 | 82 | 574 | –9.67 | –67.69 |
| 75-79 | 4 | 77 | 308 | –14.67 | –58.68 |
| 70-74 | 3 | 72 | 216 | –19.67 | –59.01 |
| | **Total= 60** | | **Total= 5500** | | **Total = 463.93** |
| | | | **Mean=91.67** | | |

**Step 1:** Identify midpoints of the class interval and mention them in column three, as can be seen above.

**Step 2:** Multiply respective frequencies and mid-points as shown in column 4.

**Step 3:** For obtained *f*X, compute mean with the help of formula M=∑X/N.

M=5500/60

=91.67

**Step 4:** x is then computed with the formula x=M–X.

**Step 5:** Average deviation is then computed with the help of following formula:

$AD = \sum |fx| / N$

$= 463.93/60$

$= 7.73$

Thus, the average deviation obtained is 7.73.

## 5.2.4 The Standard Deviation (SD)

Standard Deviation is the most stable measure of variability. Therefore, it is most commonly used in research studies. Standard deviation can be computed for ungrouped and grouped data.

### 5.2.4.1 Calculation of Standard Deviation (SD) for Ungrouped Data

Standard deviation for ungrouped data can be computed by the following formula:

$\mathbf{SD = \sqrt{\sum x^2 / N}}$

The above formula can be explained by the following example.

Let us understand the steps in computation of standard deviation for ungrouped data with the help of example given below:

| Scores (X) | Deviation from the mean (x) | Deviation square ($x^2$) |
|---|---|---|
| 52 | -8 | 64 |
| 50 | -10 | 100 |
| 56 | -4 | 16 |
| 68 | 8 | 64 |
| 65 | 5 | 25 |
| 62 | 2 | 4 |
| 57 | -3 | 9 |
| 70 | 10 | 100 |
| **Total = 480** | | $\mathbf{\sum x^2 = 382}$ |
| **Mean= 60** | | |

**Step 1:** Add all the scores ($\sum x$) and divide this sum by the number of scores (N) and find out mean. Mean of the given scores is

$M = \sum X / N$

$= 480/8$

$= 60.$

**Step 2:** Find out deviation x by computing X – x, as given in second column above.

**Step 3:** Square all the deviation to get $x^2$.

**Step 4:** Add all the squared deviation to get $\sum x^2$.

**Step 5:** Compute standard deviation with the help of the formula:

**SD=** $\sqrt{(\sum x^2)/N}$

$= \sqrt{382/8}$

$= \sqrt{47.7}$

$= 6.91$

Thus, the standard deviation for this data is 6.91.

### 5.2.4.2 Computations of SD for Grouped Data by Long Method

Standard deviation of grouped data can be computed by the formula,

**SD=** $\sqrt{\sum fx^2/N}$

Where,

$\sum fx^2$= When frequencies ($f$) are multiplied with their respective deviation squared ($x^2$), $fx^2$ is obtained. Total of all $fx^2$ is $\sum fx^2$.

N= Total number of scores

Let us understand the steps in computation of standard deviation for grouped data with the help of example given below:

| Class interval | Frequency ($f$) | Midpoint (X) | $f$X | Deviation of midpoint from the mean (x= X- M) | Deviation squared ($x^2$) | $fx^2$ |
|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| 127-129 | 1 | 128 | 128 | 17.6 | 309.76 | 309.76 |
| 124-126 | 2 | 125 | 250 | 14.6 | 213.16 | 426.32 |
| 121-124 | 2 | 122 | 244 | 11.6 | 134.56 | 269.12 |
| 118-120 | 2 | 119 | 238 | 8.6 | 73.96 | 147.92 |
| 115-117 | 4 | 116 | 464 | 5.6 | 31.36 | 125.44 |
| 112-114 | 4 | 113 | 452 | 2.6 | 6.76 | 27.04 |
| 109-111 | 4 | 110 | 440 | –0.4 | 0.16 | 0.64 |
| 106-108 | 2 | 107 | 214 | –3.4 | 11.56 | 23.12 |
| 103-106 | 2 | 104 | 208 | –6.4 | 40.96 | 81.92 |
| 100-102 | 2 | 101 | 202 | –9.4 | 88.36 | 176.72 |
| | **Total= 25** | | **∑fx= 2760** | | | **∑fx²= 1588** |

**Step 1:** Midpoint is computed for respective class intervals and entered in column 3 as shown in the above table.

**Step 2:** $f$X is then computed by multiplying the frequencies and the midpoint, the values thus obtained are entered in column 4.

**Step 3:** Add all the scores under $f$X and divide this sum by the number of scores (N) and find out mean. Thus, M= $\sum f$X/ N, that is, 2760/ 25= 110.4. The mean obtained is 110.4.

**Step 4:** x is now computed by subtracting the M from X (midpoint). The values thus obtained are entered in column 5.

**Step 5:** The x is then squared to obtain $x^2$ (column 6).

**Step 6:** $fx^2$ is then computed by multiplying $f$ and $x^2$. The values thus obtained are entered in column 7.

**Step 7:** Add the $fx^2$ to obtain $\sum fx^2$. In the present example it is obtained as 1588.

**Step 8:** Compute standard deviation with help of the formula:

**SD=** $\sqrt{\sum fx^2/N}$

$= \sqrt{1588/25}$

$= \sqrt{63.52}$

$= 7.97$

The standard deviation thus obtained is 7.97.

### 5.2.4.3 Calculation of SD for Grouped Data by Short Method

Standard deviation from grouped data can also be computed by the following formula.

**SD= i** $\sqrt{\sum fx'^2/N-(\sum(fx')/N)^2}$

where,

i= The size of the class interval

$\sum fx'^2$=When frequencies ($f$) are multiplied with their respective deviation squared ($x'^2$), $fx'^2$ is obtained. Total of all $fx'^2$ is $\sum fx'^2$.

Let us understand the steps in computation of standard deviation for grouped data by short method with the help of example given below:

| **Class interval** | **Frequency ($f$)** | **Midpoint (X)** | **Deviation of midpoint from the mean ($x'$= X- AM/i)** | **($fx'$)** | **$fx'^2$** |
|---|---|---|---|---|---|
| (1) | (2) | (3) | (5) | (6) | (7) |
| 127-129 | 1 | 128 | 4 | 4 | 16 |
| 124-126 | 2 | 125 | 3 | 6 | 36 |
| 121-124 | 2 | 122 | 2 | 4 | 16 |
| 118-120 | 2 | 119 | 1 | 2 | 4 |
| 115-117 | 4 | 116 | 0 | 0 | 0 |
| 112-114 | 4 | 113 | -1 | -4 | 16 |
| 109-111 | 4 | 110 | -2 | -8 | 64 |
| 106-108 | 2 | 107 | -3 | -6 | 36 |
| 103-106 | 2 | 104 | -4 | -8 | 64 |
| 100-102 | 2 | 101 | -5 | -10 | 100 |
| | **Total= 25** | | | **$\sum fx'$ = -20** | **$\sum fx'^2$= 352** |

**Step 1:** Find out midpoint of each class interval, that is entered in column 3.

**Step 2:** Assume one value as mean. In this example, assumed mean is taken as 116.

**Step 3:** Find out the difference between midpoint and assumed mean and divide it by class intervals to get x′, (x′= X- AM/i) and enter the obtained value in column 5.

**Step 4:** Multiply each x′ by respective frequency and get $fx'$ (column 6).

**Step 5:** $fx'$ is squared to get $fx'^2$. Add all the $fx'^2$ to obtain $\sum fx'$.

**Step 6:** Compute standard deviation with the help of the formula

**SD= i** $\sqrt{\sum fx'^2/N-(\sum(fx')/N)^2}$

$= 3\sqrt{352/25-(-20/25)^2}$

$= 3\sqrt{14.08-(-0.8)^2}$

$= 3\sqrt{14.08-0.64}$

$= 3\sqrt{3.44}$

$= 3 \times 1.85$

$= 5.55$

The standard deviation thus obtained is 5.55.

**Check Your Progress 1**

1) What is the formula for range?

2) List the steps in computation of standard deviation for ungrouped data.

......................................................................................................................

......................................................................................................................

......................................................................................................................

......................................................................................................................

......................................................................................................................

## 5.3 LET US SUM UP

In this unit, we covered the computation of different measures of variability, like the range, average deviation, quartile deviation and standard deviation. Computation of each of the measures of dispersion was discussed with the help of steps and examples.

## 5.4 REFERENCES

Garrett, H.E. (1981), *Statistics in Psychology and Education*, (Tenth edition), Bombay, Vakils Feffer and Simons Ltd.

McBride, Dawn M. (2018). *The Process of Statistical Analysis in Psychology.* Sage. USA

Minium, E.W., King, B.M. & Bear. G (2001). *Statistical Reasoning in Psychology and Education* (3rd edition), Singapore, John Wiley &Sons, Inc.

Mohanty, B. & Misra, Santa (2016). *Statistics for Behavioural and Social Sciences.* Sage. New Delhi.

## 5.5 ANSWERS TO CHECK YOUR PROGRESS

**Check Your Progress 1**

1) What is the formula for range?

   R=H-L

2) List the steps in computation of standard deviation for ungrouped data.

The steps in computation of standard deviation for ungrouped data are

**Step 1:** Add all the scores (∑x) and divide this sum by the number of scores (N) and find out mean.

**Step 2:** Find out deviation x by computing X- x.

**Step 3:** Square all the deviation to get $x^2$.

**Step 4:** Add all the squared deviation to get $\sum x^2$.

**Step 5:** Compute standard deviation with the help of the formula.

## 5.6 UNIT END QUESTIONS

1) Compute the range, average deviation and standard deviation from the following ungroupeddata:

   a) 30, 35, 36, 39, 42, 46, 38, 34, 35

   b) 52, 50, 56, 68, 65, 62, 57, 70

2) Calculate quartile deviationforthefollowingscores:

   6, 3, 9, 9, 5, 7, 9, 6, 8, 4, 8, 5, 7, 9, 3, 2, 9, 5, 7

3) Calculate Average deviation of the following scores:

| Class Interval | Frequency |
|---|---|
| 40-44 | 3 |
| 35-39 | 4 |
| 30-34 | 6 |
| 25-29 | 12 |
| 20-24 | 7 |
| 15-19 | 5 |
| 10-14 | 1 |

4) Calculate the Quartile deviation and Standard Deviation for the following frequency distribution:

| **Scores** | **Frequency** |
|---|---|
| 70-71 | 2 |
| 68-69 | 2 |
| 66-67 | 3 |
| 64-65 | 4 |
| 62-63 | 6 |
| 60-61 | 7 |
| 58-59 | 5 |
| 56-57 | 1 |
| 54-55 | 2 |
| 52-53 | 3 |
| 50-51 | 1 |

118 blank