EXERCISE 6.2

Q. 1. What do you mean by measures of central tendency?

Ans. The Specific value of the variable around which the majority of the observations of the data tend to concentrate is called average or central value. The measures or techniques that are used to determine the central value are called measure of central tendency.

Examples.

- i. Arithmetic mean
- ii. Median
- iii. Mode
- iv. Geometric mean
- v. Harmonic mean

Q.2. Define Arithmetic mean, geometric mean, Harmonic mean, mode and median.

Answer

(i) **Arithmetic Mean:**

Arithmetic mean is a measure that determines a value of the variable under study by dividing the sum of all values of the variable by their number of observations. We denote Arithmetic mean by $\overline{X}$. In symbols it is defined as.

$$\overline{X}=\frac{\sum X}{n}=\frac{\text{Sum of all values of observations}}{\text{No. of observations}}$$

(ii) Mode:

The most frequent occurring observation in the data or the observation that occurs maximum number of times in the given data is called mode

Mode = the most frequent or common value in the ungrouped data.

(iii) Median:

Median is the middle most observation in an arranged data set. It divides the data set into two equal parts. Median is represented by $\tilde{x}$. For ungrouped data median is calculated by the following formula.

- When n is odd

$$\tilde{x} = \left(\frac{n+1}{2}\right) \text{th term in arranged data}$$

- When n is even

$$\tilde{x} = \frac{1}{2}\left[\frac{n}{2}\text{th term} + \frac{n+2}{2}\text{ th term}\right]$$

(iv) **Geometric Mean:**

Geometric mean of a variable x is the n^{th} positive root of the product of the $x_1, x_2, x_3, \ldots\ldots\ldots\ldots, x_n$ observation. In symbols we write, (Basic Formula)

$$G.M = (X_1 . X_2 . X_3 . \ldots\ldots . X_n)^{1/n}$$

The Geometric Mean can also be calculated by using logarithm.

- **For Ungrouped data:**

$$G.M = Anti\log\left(\frac{\sum \log X}{n}\right)$$

- **For Grouped data:**

$$G.M = Anti\log\left(\frac{\sum f \log X}{\sum f}\right)$$

(v) **Harmonic Mean:**

Harmonic mean refers to the value obtained by reciprocating the mean of the reciprocal of $x_1, x_2, x_3, \ldots\ldots, x_n$ observations. In symbols,

- **For Ungrouped data:**

$$H.M = \frac{n}{\sum \frac{1}{X}}$$

- **For Grouped data:**

$$H.M = \frac{n}{\sum \frac{f}{X}}$$

Q.3. Find arithmetic mean by direct method for the following set of data.

(i) 12, 14, 17, 20, 24, 29, 35, 45
(ii) 200, 225, 350, 375, 270, 320, 290

(i) **Solution:**
Arithmetic mean by direct formula
X = 12, 14, 17, 20, 24, 29, 35, 45

$$\overline{X} = \frac{\sum X}{n}$$

$$\overline{X} = \frac{12+14+17+20+24+29+35+45}{8}$$

$$\overline{X} = \frac{196}{8}$$

$$\overline{X} = 24.5$$

(ii) **Solution:**
Arithmetic mean by direct formula
x = 200, 225, 350, 375, 270, 320, 290

$$\overline{X} = \frac{\sum X}{n}$$

$$\overline{X} = \frac{200+225+350+375+270+320+290}{7}$$

$$\overline{X} = \frac{2030}{7}$$

$$\overline{X} = 290$$

Q.4. For each of the data in Q. No. 3. Compute arithmetic mean using indirect method.

(i) Solution:
Arithmetic Mean by indirect (short) formula
X = 12, 14, 17, 20, 24, 29, 35, 45
Let Assumed mean = A = 20
Number of Observations = n = 8

X	D = X – A
12	12 – 20 = –8
14	14 – 20= –6
17	17 – 20 = –3
20	20 – 20 = 0
24	24 – 20 = 4
29	29 – 20 = 9
35	35 – 20 = 15
45	45 – 20 = 25
	$\sum D = 36$

$$\overline{X} = A + \frac{\sum D}{n}$$

$$\overline{X} = 20 + \frac{36}{8}$$

$$\overline{X} = 20 + 4.5$$

$$\overline{X} = 24.5$$

(ii) Solution:

Arithmetic mean by indirect (short) formula

Let assumed mean = A = 225

Number of observations = n = 7

X	D = X − A
200	200 − 225 = −25
225	225 − 225 = 0
270	270 − 225 = 45
290	290 − 225 = 65
320	320 − 225 = 95
350	350 − 225 = 125
375	375 − 225 = 150
	$\sum D = 455$

$$\overline{X} = A + \frac{\sum D}{n}$$

$$\overline{X} = 225 + \frac{455}{7}$$

$$\overline{X} = 225 + 65$$

$$\overline{X} = 290$$

Q.5. The marks obtained by students of Class XI in mathematics are given below. Compute arithmetic mean by direct and indirect methods.

Classes /Groups	Frequency
0 – 9	2
10 – 19	10
20 – 29	5
30 – 39	9
40 – 49	6
50 – 59	7
60 – 69	1

Solution:

(i) Arithmetic mean by Direct Method

Classes / Groups	f	Class Marks or Mid Points (x)	fx
0 – 9	2	4.5	9
10 – 19	10	14.5	145
20 – 29	5	24.5	122.5
30 – 39	9	34.5	310.5
40 – 49	6	44.5	267
50 – 59	7	54.5	381.5
60 – 69	1	64.5	64.5
Total	$\sum f = 40$		$\sum fx = 1300$

$$\overline{X} = \frac{\sum fx}{\sum f} = \frac{1300}{40}$$

$$\overline{X} = 32.5$$

$\overline{X}$ = 32.5 Marks

(ii) Arithmetic mean by indirect (short) method

Let assumed mean = A= 34.5

Class /Groups	f	Class Mark (x)	D= X–A	fD
0 – 9	2	4.5	4.5 – 34.5 = –30	–60
10 – 19	10	14.5	14.5 – 34.5 = –20	–200
20 – 29	5	24.5	24.5 – 34.5= –10	–50
30 – 39	9	34.5	34.5 – 34.5 = 0	0
40 – 49	6	44.5	44.5 – 34.5 = 10	60
50 – 59	7	54.5	54.5 – 34.5 = 20	140
60 – 69	1	64.5	64.5 – 34.5 = 30	30
Total	$\sum f = 40$		$\sum fD = -80$	

$$\overline{X} = A + \frac{\sum fD}{\sum f}$$

$$\overline{X} = 34.5 + \frac{-80}{40}$$

$$\overline{X} = 34.5 + (-2)$$

$$\overline{X} = 34.5 - 2$$

$\overline{X}$ = 32.5 Marks

Q.6. The following data relates to the ages of children in a school. Compute the mean age by direct and short cut method taking any provincial mean. (Hint: Take A= 8)

Class Limits	Frequency
4 – 6	10
7 – 9	20
10 – 12	13
13 – 15	7
Total	50

Also compute Geometric mean and Harmonic mean

Solution:

(i) **Arithmetic mean by direct formula.**

Class Limits	Frequency (f)	Class Marks (x)	fx
4 – 6	10	5	50
7 – 9	20	8	160
10 – 12	13	11	143
13 – 15	7	14	98
Total	$\sum f = 50$		$\sum fx = 451$

$$\overline{X} = \frac{\sum fx}{\sum f} = \frac{451}{50}$$

$$\overline{X} = 9.02$$

(ii) **Arithmetic Mean by Indirect (Short) method**

Here A = 8 Assumed mean (A) = 8

Class Limits	f	Class Marks (x)	D= x–A	fD
4 – 6	10	5	5 – 8= –3	– 30
7 – 9	20	8	8 – 8= 0	0
10 – 12	13	11	11 – 8=3	39
13 – 15	7	14	14 – 8=6	42
Total	50			$\sum fD - 51$

$$\overline{X} = A + \frac{\sum fD}{\sum f} \Rightarrow \overline{X} = 8 + \frac{51}{50}$$

$$\overline{X} = 8 + 1.02 \Rightarrow \overline{X} = 9.02$$

(iii) **Geometric Mean:**

Class Limit	f	Marks (x)	log x	f log x
4 – 6	10	5	0.6990	6.990
7 – 9	20	8	0.9031	18.062
10 – 12	13	11	1.0414	13.5382
13 – 15	7	14	1.1461	8.0227
Total	50		$\sum f \log x = 46.6129$	

$$G.M = \text{Anti log}\left(\frac{\sum f \log x}{\sum f}\right)$$

$$G.M = \text{Anti log}\left(\frac{46.6129}{50}\right)$$

$$G.M = \text{Anti log}(0.932258)$$

or $$G.M = \text{Anti log}(0.9323)$$

$$G.M = 8.557$$

(iv) **Harmonic Mean:**

Class Limits	f	Class Marks (x)	f/x
4–6	10	5	2
7–9	20	8	2.5
10–12	13	11	1.1818
13–15	7	14	0.5
Total	$\sum f = 50$		$\sum f/x = 6.1818$

$$H.M = \frac{\sum f}{\sum\left(\frac{f}{x}\right)}$$

$$H.M = \frac{50}{6.1818} = 8.088$$

or H.M = 8.089

Q. 7 The following data shows the number of children in various families. Find mode and median.

9, 11, 4, 5, 6, 8, 4, 3, 7, 8, 5, 5, 8, 3, 4, 9, 12, 8, 9, 10, 6, 7, 7, 11, 4, 4, 8, 4, 3, 2, 7, 9, 10, 9, 7, 6, 9, 5

Solution:

Arranged Data:

2, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 9, 9, 9, 9, 9, 9, 10, 10, 11, 11, 12

Mode:

From the arranged data we see that 4 and 9 both are repeated 6 times i.e. the most frequent values are 4 and 9. So,

Mode = 4 and 9

Median:

As n is even (n=38). So,

$$\text{Median} = \frac{1}{2}\left[\frac{n}{2}\text{th item} + \frac{n+2}{2}\text{th item}\right]$$

$$\text{Median} = \frac{1}{2}\left[\frac{38}{2}\text{th item} + \frac{38+2}{2}\text{th item}\right]$$

$$\text{Median} = \frac{1}{2}[\text{19th item} + \text{20th item}]$$

$$\text{Median} = \frac{1}{2}[7+7]$$

$$\text{Median} = \frac{1}{2}[14]$$

$$\text{Median} = 7$$

Q. 8 Find Modal number of heads for the following distribution showing the number of heads when 5 coins are tossed. Also determine median.

X (number of heads)	Frequency (f)
1	3
2	8
3	5
4	3
5	1

Solution: we first make cumulative frequency column as given below:

X	Frequency	Cumulative frequency
1	3	3
2	8	3 + 8 = 11
3	5	11 + 5 =16
4	3	16 + 3 =19
5	1	19 + 1 = 20

(i) Mode: As most frequent number is 2 i.e., 2 is repeated 8 times. So, Mode = 2 Ans.

(ii) Median:

Median = the class containing $\left[\frac{n}{2}\right]^{th}$ observation

Median = the class containing $\left[\frac{20}{2}\right]^{th}$ observation

Median = the class containing $(10)^{th}$ observation

Median = 2

Q. 9 The following frequency distribution shows the weights of boys in kilogram. Compute mean, median, mode.

Class Interval	Frequency
1 – 3	2
4 – 6	3
7 – 9	5
10 – 12	4
13 – 15	6
16 – 18	2
19 – 21	1

Solution:

(i) Arithmetic Mean

Class Intervals	Frequency (f)	Class marks (x)	fx
1 – 3	2	2	4
4 – 6	3	5	15
7 – 9	5	8	40
10 – 12	4	11	44
13 – 15	6	14	84
16 – 18	2	17	34
19 – 21	1	20	20
	$\Sigma f = 23$		$\Sigma fx = 241$

$$\overline{X} = \frac{\Sigma fx}{\Sigma f} = \frac{241}{23} = 10.478kg$$

(ii) Median:

Class Intervals	(f)	Class Boundaries	Cumulative Frequency
1 – 3	2	0.5 – 3.5	2
4 – 6	3	3.5 – 6.5	2 + 3 = 5
7 – 9	5	6.5 – 9.5	5 + 5 = 10
10 – 12	**4**	**9.5 – 12.5**	**10 + 4 = 14**
13 – 15	6	12.5 – 15.5	14 + 6 = 20
16 – 18	2	15.5 – 18.5	20 + 2 = 22
19 – 21	1	18.5 – 21.5	22 + 1 = 23

Median = class containing

$$\left(\frac{n}{2}\right)^{th} = \left(\frac{23}{2}\right)^{th} = 11.5^{th} \text{ item}$$

As 11.5^{th} item is present in the group (9.5 – 12.5) so median group is (9.5 – 12.5).

$$\text{Median} = l + \frac{h}{f}\left(\frac{n}{2} - c\right)$$

$$\text{Median} = 9.5 + \frac{3}{4}\left(\frac{23}{2} - 10\right)$$

$$\text{Median} = 9.5 + \frac{3}{4}(11.5 - 10)$$

$$\text{Median} = 9.5 + \frac{3}{4}(1.5)$$

$$\text{Median} = 9.5 + \frac{4.5}{4}$$

$$\text{Median} = 9.5 + 1.125$$

$$\text{Median} = 10.625 \text{ kg}$$

(iii) Mode:

Class Intervals	Frequency	Class Boundaries
1 – 3	2	0.5 – 3.5
4 – 6	3	3.5 – 6.5
7 – 9	5	6.5 – 9.5
10 – 12	$f_1 \to$ 4	9.5 – 12.5
13 – 15	$f_m \to$ 6	12.5 – 15.5
16 – 18	$f_2 \to$ 2	15.5 – 18.5
19 – 21	1	18.5 – 21.5

As the group (12.5–15.5) has maximum frequency (i.e. 6) so modal group is (12.5–15.5).

$$\text{Mode} = l + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h$$

$$\text{Mode} = 12.5 + \frac{(6-4)}{2(6)-4-2} \times 3$$

$$\text{Mode} = 12.5 + \frac{(2\times3)}{12-6}$$

$$\text{Mode} = 12.5 + \frac{6}{6}$$

$$\text{Mode} = 12.5 + 1$$

$$\text{Mode} = 13.5 \text{ kg}$$

Q.10 A student obtained the following marks at a certain examination: English 73, Urdu 82, Mathematics 80, History 67 and Science 62.

(i) If the weights accorded these marks are 4, 3, 3, 2 and 2, respectively, what is an appropriate average mark?

(ii) What is the average mark if equal weights are used?

Solution:

Subjects	Marks (x)	Weights (w)	Wx
English	73	4	292
Urdu	82	3	246
Maths	80	3	240
History	67	2	134
Science	62	2	124
		$\sum w = 14$	$\sum wx = 1036$

$$\text{Weighted mean } \overline{X}_w = \frac{\sum wx}{\sum w}$$

$$\overline{X}_w = \frac{1036}{14}$$

$$\overline{X}_w = 74 \text{ marks}$$

(ii) Solution: Let equal weight is 2, then

Subjects	Marks	Weight	Wx
English	73	2	146
Urdu	82	2	164
Maths	80	2	160
History	67	2	134
Science	62	2	124
		$\sum w = 10$	$\sum wx = 728$

$$\text{Equal weighted mean} = \frac{\sum wx}{\sum w} = \frac{728}{10}$$

$$= 72.8 \text{ marks}$$

Q.11 On a vacation trip a family bought 21.3 liters of petrol at 39.90 rupees per liter, 18.7 liters at 42.90 rupees per liter, and 23.5 liters at 40.90 rupees per liter. Find the mean price paid per liter.

Solution:

Liters (w)	Price per liters (x)	Payments (Rs) (wx)
21.3	39.90	849.87
18.7	42.90	802.23
23.5	40.90	961.15
Total liters $\Sigma w = 63.5$		**Total Payments $\Sigma wx = 2613.25$**

$$\text{Average price} = \frac{\text{Total payment}}{\text{Total liters}}$$

$$\overline{X}_w = \frac{\sum wx}{\sum w}$$

$$\overline{X}_w = \frac{2613.25}{63.5}$$

$$\overline{X}_w = 41.15 \text{ rupees per liter}$$

Q. 12 Calculate simple moving average of 3 year from the following data.

Years	Values
2001	102
2002	108
2003	130
2004	140
2005	158
2006	180
2007	196
2008	210
2009	220
2010	230

Solution:

Years	Values	3 years Moving	
		Total	Average
2001	102	–	–
2002	108	340	$\frac{340}{3}=113.33$
2003	130	378	$\frac{378}{3}=126$
2004	140	428	$\frac{428}{3}=142.66$
2005	158	478	$\frac{478}{3}=159.33$
2006	180	534	$\frac{534}{3}=178$
2007	196	586	$\frac{586}{3}=195.33$
2008	210	626	$\frac{626}{3}=208.67$
2009	220	660	$\frac{660}{3}=220$
2010	230	–	–

Q.13 Determine graphically for the following data and check your answer by using formulae.

(i) Median and Quartiles using cumulative frequency polygon.

(ii) Mode using Histogram

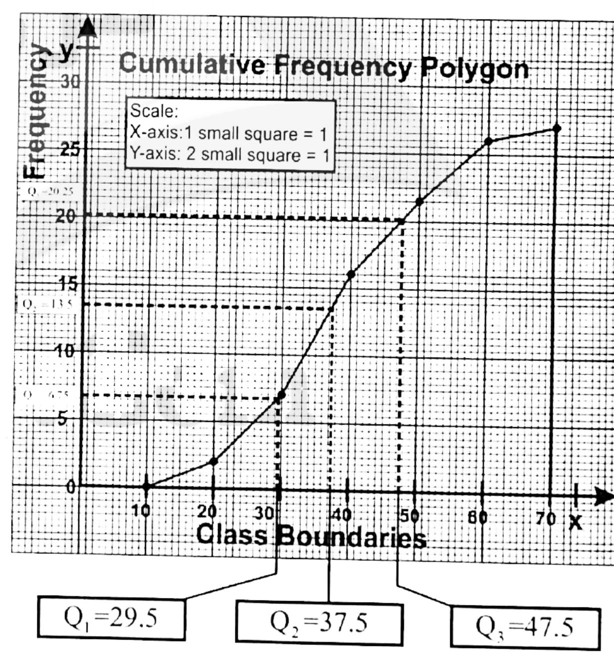
Class Boundaries	Frequency
10 – 20	2
20 – 30	5
30 – 40	9
40 – 50	6
50 – 60	4
60 – 70	1

Solution:

(i) Less than Frequency Table:

Class Boundaries	C. F
Less than 10	0
Less than 20	2
Less than 30	7
Less than 40	16
Less than 50	22
Less than 60	26
Less than 70	27

Cumulative Frequency Polygon:



Finding Q_1: Take the value of Q_1 on the x-axis against the $\frac{27}{4}$th observation (6.75) on the y-axis which is 29.5.

Finding Q_2: Take the value of Q_2 on the x-axis against the $2\left(\frac{27}{4}\right)$th observation (13.5) on the y-axis which is 37.5.

Finding Q_3: Take the value of Q_3 on the x-axis against the $3\left(\frac{27}{4}\right)$th observation (20.25) on the y-axis which is 47.5.

Verification by Formulae:

Class Boundaries	f	C. F
10–20	2	2
20–30	5	2+5=7
30–40	9	7+9=16
40–50	6	16+6=22
50–60	4	22+4=26
60–70	1	26+1=27

Verification of Q_1 by Formula:

Q_1 Class = Class Containing $(\frac{n}{4})$th item

Q_1 Class = Class Containing $(\frac{27}{4})$th item

Q_1 Class = Class Containing (6.75)th item

As 6.75 item is present in the group (20–30) so Q_1 Class is (20–30)

$$\text{Value of } Q_1 = l + \frac{h}{f}(\frac{n}{4} - c)$$

$$= 20 + \frac{10}{5}\left(\frac{27}{4} - 2\right)$$

$$= 20 + \frac{10}{5}(6.75 - 2)$$

$$= 20 + \frac{10}{5}(4.75)$$

$$= 20 + \frac{47.5}{5}$$

$$= 20 + 9.5$$

$$Q_1 = 29.5$$

This result (29.5) is exactly same as obtained from the graph.

Verification of Median (Q_2) by Formula:

Median Class = Class Containing $2\left(\frac{n}{4}\right)$th item

Median Class = Class Containing $(\frac{27}{2})$th item

Median Class = Class Containing (13.5)th item

As 13.5 item is present in the group (30–40) so, Median Group is (30–40)

$$\text{Median} = l + \frac{h}{f}(\frac{n}{2} - c)$$

$$= 30 + \frac{10}{9}\left(\frac{27}{2} - 7\right)$$

$$= 30 + \frac{10}{9}(13.5 - 7)$$

$$= 30 + \frac{10}{9}(6.5)$$

$$= 30 + \frac{65}{9}$$

$$= 30 + 7.22 \Rightarrow Q_2 = 37.22$$

The result is very close to the value (37.5) which is obtained from graph.

Verification of Q_3 by Formula:

Q_3 Class = Class Containing $3(\frac{n}{4})$th item

Q_3 Class = Class Containing $3(\frac{27}{4})$th item

Q_3 Class = Class Containing 3(6.75)th item

Q_3 Class = Class containing 20.25^{th} item

As 20.25^{th} item is present in the group (40–50) so, Q_3 Class is (40–50).

$$\text{Value of } Q_3 = l + \frac{h}{f}(\frac{3n}{4} - c)$$

$$= 40 + \frac{10}{6}\left(\frac{3 \times 27}{4} - 16\right)$$

$$= 40 + \frac{10}{6}(20.25 - 16)$$

$$= 40 + \frac{10}{6}(4.25)$$

$$= 40 + \frac{42.5}{6}$$

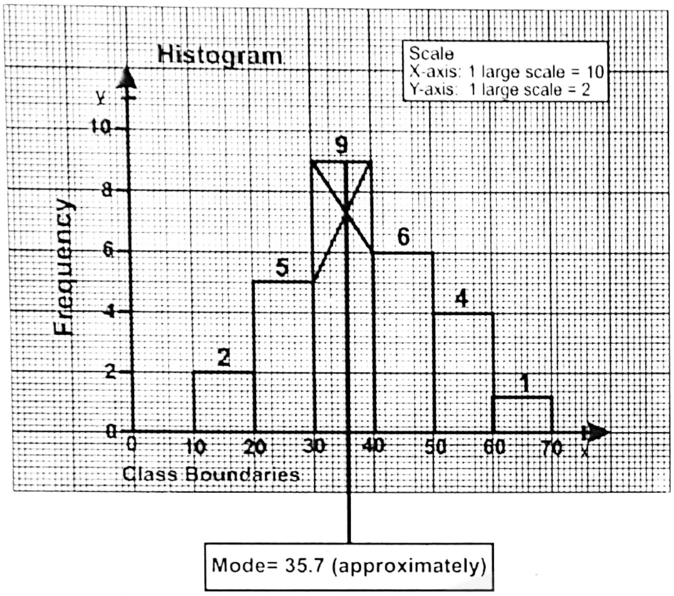
$$= 40 + 7.08 \Rightarrow Q_3 = 47.08$$

The result is very close to the value (47.5) which is obtained from graph.

(ii) Solution:

Class Boundaries	Frequency
10–20	2
20–30	5
30–40	9
40–50	6
50–60	4
60–70	1

Histogram:



From the Graph

Mode = 35.75

Verification of Mode by Formula

Class Boundaries	Frequency
10–20	2
20–30	$f_1 \rightarrow$ 5
30–40	$f_m \rightarrow$ 9
40–50	$f_2 \rightarrow$ 6
50–60	4
60–70	1

As the group (30–40) has maximum frequency (9).

So modal group is (30–40)

$$\text{Mode} = l + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h$$

$$= 30 + \frac{(9-5)\times 10}{2(9)-5-6}$$

$$= 30 + \frac{4\times 10}{18-11}$$

$$= 30 + \frac{40}{7}$$

$$= 30 + 5.71$$

Mode = 35.71

This result is very close to the value (35.5) which is obtained from the graph.

Dispersion

Statistically, dispersion means the spread or scatterness of observations in a data set.

The spread or scatterness in a data set can be seen in two ways.

i. The spread between two extreme observations in a data set.

ii. The spread of observations around an average say their arithmetic mean.

Measures of Dispersion

The measures or techniques that are used to determine the degree or extent of variation in a data set are called Measures of Dispersion.

(1) Range:

Range measures the extent of variation between two extreme observations of a data set. It is given by the formula.

Range = $X_{max} - X_{min} = X_m - X_o$ Where,

$X_{max} = X_m$: the maximum observation.

$X_{min} = X_o$: the minimum observation.

Note: The formula to find range for grouped continuous data is given below.

Range = (upper class boundary of last group)− (lower class boundary of first group)

Example 1: Find Range for the following weights of students.

110,109,84,89,77,104,74,97,49,59,103,62.

Solution: Given that,

Maximum Value = X_m = 110

Minimum Value = X_o = 49

$R = X_m - X_o$

$R = 110 - 49 = 61$

Example 2: Find the Range for the following distribution.

Class / groups	f
10–19	10
20–29	5
30–39	9
40–49	6
50–59	7
60–69	1
Total	$\sum f = 38$

Solution: We find class boundaries for the given data as follows.

Class / groups	Class Boundaries	f
10–19	9.5–19.5	10
20–29	19.5–29.5	5
30–39	29.5–39.5	9
40–49	39.5–49.5	6
50–59	49.5–59.5	7
60–69	59.5–69.5	1
Total		$\sum f = 38$

Upper class boundary of last group = 69.5
Lower class boundary of first group = 9.5
Range = 69.5–9.5 = 60

(2) **Variance:**

Variance is defined as the mean of the squared deviation of X_i (i = 1,2,……., n) observations from their arithmetic mean. Variance is denoted by S^2.

In symbols.

$$\text{Variance of X=var (X)} = S^2 = \frac{\sum(X-\overline{X})^2}{n}$$

(3) **Standard Deviation:**

Standard deviation is defined as the positive square root of mean of the squared deviations of X_i (i= 1,2,….., n) observations from their arithmetic mean. It is denoted by "S" and it is abbreviated as "S.D".

In symbols Standard Deviation of X is

$$\text{S.D (X)} = S = \sqrt{\frac{\sum(X-\overline{X})^2}{n}}$$

Computation of variance and standard deviation

Ungrouped Data:

The formulae of variance are given by:

- Definitional Formula

$$\text{Var (X)} = S^2 = \frac{\sum(X-\overline{X})^2}{n}$$

- Computational Formula

$$\text{Var (X)} = S^2 = \frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2$$

The formulae of standard deviation are given by:

- Definitional Formula

$$\text{S.D (X)} = S = \sqrt{\frac{\sum(X-\overline{X})^2}{n}}$$

- Computational Formula

$$\text{S.D (X)} = S = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2}$$

Example 3: The marks of six students in Mathematics are as follows. Determine variance and standard deviation. Also interpret the results.

Student No.	1	2	3	4	5	6
Marks	60	70	30	90	80	42

Solution: Let X = marks of a student.

X	X^2	$X-\overline{X}$	$(X-\overline{X})^2$
60	3600	–2	4
70	4900	8	64
30	900	–32	1024
90	8100	28	784
80	6400	18	324
42	1764	–20	400
$\sum X = 372$	$\sum X^2 = 25664$		$\sum(X-\overline{X})^2 = 2600$

Mean: $\overline{X} = \frac{\sum X}{n}$, $\overline{X} = \frac{372}{6} = 62$ marks

Variance:

- By Definitional Formula

$$\text{Var (X)} = S^2 = \frac{\sum(X-\overline{X})^2}{n}$$

$$S^2 = \frac{2600}{6} = 433.33 \text{ (square marks)}$$

- By Computational Formula

$$\text{Var (X)} = S^2 = \frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2$$

$$S^2 = \frac{25664}{6} - \left(\frac{372}{6}\right)^2 = 42773.333 - 3844$$

$$S^2 = 433.333 \text{(square marks)}$$

Standard Deviation:

- By Definitional Formula

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

$$S = \sqrt{\frac{2600}{6}} = \sqrt{433.33} = 20.81 \text{marks}$$

- By Computational Formula

$$S.D(X) = S = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2}$$

$$S = \sqrt{\frac{25664}{6} - \left(\frac{372}{6}\right)^2} = \sqrt{42773.333-3844}$$

$$S = \sqrt{433.333} = 20.81 \text{ marks}$$

Grouped Data:

Following Formulae are used to find variance and standard deviation

Variance:

- Definitional Formula

$$Var(x) = S^2 = \frac{\sum f(x-\bar{x})^2}{\sum f}$$

- Computational Formula

$$Var(x) = S^2 = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$$

Standard Deviation:

- Definitional Formula

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

- Computational Formula

$$S.D(x) = S = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

Example 4: For the following data showing weights of toffee boxes in gm. determine the variance and standard deviation by using direct methods.

X (gm)	f
4.5	2
14.5	10
24.5	5
34.5	9
44.5	6
54.5	7
64.5	1

Solution:

X (gm)	f	fx	$x-\bar{x}$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
4.5	2	9	–28	784	1568
14.5	10	145	–18	324	3240
24.5	5	122.5	–8	64	320
34.5	9	310.5	2	4	36
44.5	6	267	12	144	864
54.5	7	381.5	22	484	3388
64.5	1	64.5	32	1024	1024
$\sum f = 40$		$\sum fx = 1300$		2600	10440

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{1300}{40} = 32.5$$

Variance:

- Using Definitional formula:

$$S^2 = \frac{\sum f(x-\bar{x})^2}{\sum f}$$

$$S^2 = \frac{10440}{40} = 261 \text{ sq.gm}$$

- Using computational formula:

$$Var(x) = S^2 = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$$

$$S^2 = \frac{52690}{40} - \left(\frac{1300}{40}\right)^2$$

$$S^2 = 1317.25 - (32.5)^2$$

$$S^2 = 1317.25 - 1056.25$$

$$S^2 = 261 \text{ sq.gm}$$

Standard Deviation

- Using Definitional formula:

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

$$S = \sqrt{\frac{10440}{40}} = \sqrt{261} = 16.155 \text{ gm}$$

- Using computational formula:

$$S.D(x) = S = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

$$S = \sqrt{\frac{10440}{40} - \left(\frac{1300}{40}\right)^2} = \sqrt{1317.25 - (32.5)^2}$$

$$S = \sqrt{1317.25 - 1056.25} = \sqrt{261} = 16.155 \text{ gm}$$

Example 5: Compare the variation about mean for the two groups of students who obtained the following marks in Statistics.

X= Marks (Section A)	Y= Marks (Section B)
60	62
70	62
30	65
90	68
80	67
40	48

Solution: In order to compare variation about mean we compute standard deviation for the two groups as follows.

X	Y	$X-\overline{X}$	$(X-\overline{X})^2$	$Y-\overline{Y}$	$(Y-\overline{Y})^2$
60	62	–2	4	0	0
70	62	8	64	0	0
30	65	–32	1024	3	9
90	68	28	784	6	36
80	67	18	324	5	25
40	48	–22	484	–14	196
372	**372**	**0**	**2684**	**0**	**266**

$$\text{Mean for group A} = \overline{x} = \frac{\sum x}{n} = \frac{372}{6} = 62$$

$$\text{Mean for group B} = \overline{y} = \frac{\sum y}{n} = \frac{372}{6} = 62$$

$$S.D(x) = \sqrt{\frac{\sum (x-\overline{x})^2}{n}} = \sqrt{\frac{2684}{6}} = \sqrt{447.333} = 21.15 \text{ gm}$$

$$S.D(y) = \sqrt{\frac{\sum (y-\overline{y})^2}{n}} = \sqrt{\frac{266}{6}} = \sqrt{44.333} = 6.66 \text{ gm.}$$

Comparison:

From the results we conclude that the variation in Group B is smaller than that of Group A. This implies the marks of students in Group B are closer to their Mean than that of Group A.