Chapter:6

Measures of variation (or) Dispersion

INTRODUCTION

The measures of central tendency (or Averages) gives us an idea of the concentration of the observations about the central part of the distribution. But these measures are inadequate to give us a complete idea of the distribution. They must be supplemented and supported by some other measures, such as dispersion.

The measurement of the scatter of the given data about the average is said to be a measure of dispersion or scatter.

1. According to Simpson and Kafka: "The measurement of the scatteredness of the mass of figures in a series about an average is called measure of variation (or) dispersion.
2. According to Spiegel: "The degree to which numerical data tend to spread about an average value is called the variation or dispersion of the data".
3. According to A.L. Bowley: "Dispersion is a measure of variation of the items".

The term dispersion is used in two senses. The first relates to the limits within which the data falls and the second takes into account the amount, absolute or relative, by which the values of the items differ from an average.

The absolute measures can be divided in the following four positional measures.

1. Range.
2. Quartile deviation.
3. Mean deviation.
4. Standard deviation.

The relative measures in each of the above four cases are called the coefficient of the respective measures such as coefficient of standard deviation, etc. The relative measures are used only for the purpose of comparison between two or more series with varying size or number of items or varying central or varying units of calculations.

CHARACTERISTICS OF A GOOD MEASURE OF DISPERSION

According to G.U. Yule, a good measure of dispersion should have the following.

1. It should be easy to understand and calculate.
2. A good measure of dispersion should be rigidly defined.
3. It should be based on all observations.
4. It should be capable of further algebraic treatment.
5. It should not be unduly affected by extreme items.
6. It should have sampling stability.
7. It should be readily comprehensible.
8. Its computation procedure should be simple.

1. Range

Range is the simplest measure of dispersion. It is the difference between the value of maximum and the value of the minimum of the distribution.

$$\text{Range} = \text{Maximum value} - \text{Minimum value}$$

$$\text{Coefficient of range} = \frac{\text{Maximum value} - \text{Minimum value}}{\text{Maximum value} + \text{Minimum value}}$$

Merits

1. It is easy to understand.
2. It is also simple to calculate.
3. It takes minimum time to calculate.

Demerits

1. It is highly affected by extreme values.
2. It cannot be calculated from frequency distribution with open end classes.
3. It does not depend on all observations and is based on only the maximum and minimum value among them.

Uses

1. It plays an important role in preparing control charts in the methods of statistical quality control.
2. It is preferably used in determining the difference in maximum and minimum temperature for predicting the variation of temperature in a day.
3. It is useful for studying the variations in prices of shares.

Individual series

EXAMPLE 6.1: Find the range and coefficient of range of the weights of 10 students from the following data:

41, 20, 15, 65, 73, 84, 53, 35, 71, 55.

Solution: Arranging the data in the ascending order, we get

15, 20, 35, 41, 53, 55, 65, 71, 73, 84.

Maximum value = 84

Minimum value = 15

$$\text{Range} = \text{Maximum value} - \text{Minimum value}$$

$$\text{Range} = 84 - 15 = 69$$

$$\text{Coefficient of range} = \frac{\text{Maximum value} - \text{Minimum value}}{\text{Maximum value} + \text{Minimum value}} = \frac{84-15}{84+15} = 0.696$$

$$\text{Coefficient of range} = 0.696$$

Discrete series

EXAMPLE 6.2: From the following data calculate the range and coefficient of range.

x	11	18	29	33	37	39	40	42	43
f	100	105	91	82	61	32	70	88	67

Solution: By arranging the data in the ascending order, we get

x	11	18	29	33	37	39	40	42	43
f	100	105	91	82	61	32	70	88	67

Maximum value of the x variable is 43
Minimum value of the x variable is 11

$$\text{Range} = \text{Maximum value} - \text{Minimum value}$$

$$\text{Range} = 43 - 11 = 32$$

$$\text{Coefficient of range} = \frac{\text{Maximum value} - \text{Minimum value}}{\text{Maximum value} + \text{Minimum value}} = \frac{43-11}{43+11} = 0.5926$$

$$\text{Coefficient of range} = 0.5926$$

Continuous series

EXAMPLE 6.3: From the data given below calculate the coefficient of range.

Marks	10–20	20–30	30–40	40–50	50–60
No. of students	22	28	44	43	49

Solution:

Marks	10–20	20–30	30–40	40–50	50–60
No. of students	22	28	44	43	49

In case of continuous series, range can be determined to find difference between the upper limit of the highest class and the lowest limit of the lowest class.

Here maximum value is 60 and minimum value is 10.

$$\text{Range} = \text{Maximum value} - \text{Minimum value}$$
$$\text{Range} = 60 - 10 = 50$$

$$\text{Coefficient of range} = \frac{\text{Maximum value} - \text{Minimum value}}{\text{Maximum value} + \text{Minimum value}} = \frac{60-10}{60+10} = 0.7143$$

$$\text{Coefficient of range} = 0.7143$$

2. Quartile Deviation or Semi-Interquartile Range

Quartile deviation is a measure of dispersion based on the upper quartile (Q_3) and lower quartile (Q_1) of a series. It is half of the difference between the upper quartile and the lower quartile. The difference is the range between these two quartiles and is called interquartile range. Half of this range is semi-interquartile range.

$$\text{Quartile deviation (QD)} = \frac{Q_3 - Q_1}{2}$$

where Q_3 = Third quartile or upper quartile

Q_1 = First quartile or lower quartile

Coefficient of quartile deviation

Quartile deviation is an absolute measure of dispersion. Its relative measure is the coefficient of quartile deviation.

$$\text{Coefficient of quartile deviation} = \frac{\frac{Q_3 - Q_1}{2}}{\frac{Q_3 + Q_1}{2}} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Merits

1. It is easy to calculate.
2. It can be easily understood.
3. It is not affected by the extreme values.
4. It has a special utility in measuring variation in case of frequency distribution with open end classes at both ends.

Demerits

1. It is not based on all the observations.
2. It is not the representative value of data.
3. It is most capable of algebraic treatments.
4. It is affected by sampling fluctuations.
5. It cannot be regarded as measure of dispersion as it really does not show the scatter around an average but rather a distance on a scale.

Uses

1. It can be used only for descriptive statistics.

EXAMPLE 6.4: From the following data calculate quartile deviation and coefficient of quartile deviation.

X: 8, 14, 59, 39, 64, 30, 26, 28, 22, 10, 44.

Solution: Arrange the series in ascending order

X: 8, 10, 14, 22, 26, 28, 30, 38, 44, 59, 64.

n = Number of values = 11

$$Q_1 = \text{Size of } \left(\frac{n+1}{4}\right)^{\text{th}} \text{ item value of the ascending order}$$

$$Q_1 = \text{Size of } \left(\frac{11+1}{4}\right)^{\text{th}} \text{ item value of the ascending order}$$

Q_1 = Size of 3rd value of the ascending order is 14

$Q_1 = 14$

$$Q_3 = \text{Size of } 3\left(\frac{n+1}{4}\right)^{\text{th}} \text{ item value of the ascending order}$$

$$Q_3 = \text{Size of } 3\left(\frac{11+1}{4}\right)^{\text{th}} \text{ item value of the ascending order}$$

Q_3 = Size of 9th value of the ascending order is 44

$Q_3 = 44$

Now interquartile range = $Q_3 - Q_1 = 44 - 14 = 30$

Semi-interquartile range or quartile deviation is

$$\text{QD} = \frac{Q_3 - Q_1}{2} = \frac{44-14}{2} = \frac{30}{2} = 15$$

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{44-14}{44+14} = \frac{30}{58} = 0.5172$$

Discrete series:

EXAMPLE 6.5: Calculate quartile deviation and coefficient of quartile deviation.

x	4	8	12	16	20	24	28	32
f	4	9	17	40	53	37	24	16

Solution:

x	f	cf
4	4	$0 + 4 = 4$
8	9	$9 + 4 = 13$
12	17	$17 + 13 = 30$
16	40	$40 + 30 = 70$
20	53	$53 + 70 = 123$
24	37	$37 + 123 = 160$
28	24	$24 + 160 = 184$
32	16	$16 + 184 = 200$
	200	

Here total frequency, $N = \Sigma f = 200$

Q_1 = Size of $\left(\frac{N+1}{4}\right)^{\text{th}}$ term value in cumulative frequency corresponding to variable x.

Q_1 = Size of $\left(\frac{200+1}{4}\right)=\left(\frac{201}{4}\right)$ = 50.25th term value in cumulative frequency corresponding to variable x is 16

$Q_1 = 16$

Q_3 = Size of $3\left(\frac{N+1}{4}\right)^{\text{th}}$ term value in cumulative frequency corresponding to variable x

Q_3 = Size of $3\left(\frac{200+1}{4}\right) = 3\left(\frac{201}{4}\right) = (3 \times 50.25)^{\text{th}}$ = 150.75th term value in cumulative frequency corresponding to variable x is 24

$Q_3 = 24$

Now interquartile range = $Q_3 - Q_1 = 24 - 16 = 8$
Semi-interquartile range or quartile deviation is

$$\text{QD} = \frac{Q_3 - Q_1}{2} = \frac{24-16}{2} = \frac{8}{2} = 4$$

Coefficient of quartile deviation= $\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{24-16}{24+16} = \frac{8}{40} = 0.2$

Continuous series

EXAMPLE 6.6: Compute quartile deviation and coefficient of quartile deviation for the following data.

Class interval	0–10	10–20	20–30	30–40	40–50	50–60	60–70
Frequency	8	16	22	30	24	12	6

Solution:

Class interval	*frequency*	*Cumulative frequency*
0–10	8	0 + 8 = 8
10–20	16	16 + 8 = 24
20–30	22	22 + 24 = 46
30–40	30	30 + 46 = 76
40–50	24	24 + 76 = 100
50–60	12	12 + 100 = 112
60–70	6	6 + 112 = 118
	118	

$$Q_1 = l_1 + \frac{\frac{N}{4} - m_1}{f_1} \times C$$

where $N = \Sigma f$ = Total frequency = 118

$$\frac{N}{4} = \frac{118}{4} = 29.5$$

m_1 = Cumulative frequency of the class = 24

f_1 = Frequency of the class = 22

C = Width of the class interval = 10

l = Lower limit of the class = $\frac{20+20}{2} = 20$

$$Q_1 = 20 + \frac{29.5 - 24}{22} \times 10 = 20 + \frac{55}{22} = 20 + 2.5 = 22.5$$

$$Q_1 = 22.5$$

$$Q_3 = l_3 + \frac{\frac{3N}{4} - m_3}{f_3} \times C$$

where $N = \Sigma f = Total\ frequency = 118$

$$\frac{3N}{4} = \frac{3 \times 118}{4} = \frac{354}{4} = 88.5$$

m_3 = Cumulative frequency of the class = 76

f_3 = Frequency of the class = 24

C = Width of the class interval = 10

l = Lower limit of the class = $\frac{40 + 40}{2} = 40$

$$Q_3 = 40 + \frac{88.5 - 76}{24} \times 10 = 40 + \frac{12.5}{24} = 40 + 5.2 = 45.2$$

$$Q_3 = 45.2$$

w interquartile range = $Q_3 - Q_1 = 45.2 - 22.5 = 22.7$

ni-interquartile range or quartile deviation is

$$\text{QD} = \frac{Q_3 - Q_1}{2} = \frac{45.2 - 22.5}{2} = \frac{22.7}{2} = 11.35$$

fficient of quartile deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{45.2 - 22.5}{45.2 + 22.5} = \frac{22.7}{67.7} = 0.3353$

3. Mean Deviation

Mean deviation of a set of observations of a series is the arithmetic mean of all the deviations, without their algebraic signs, taken from its central value (mean or median or mode). In other words, it is in the average of the modulus of the deviations of the observations in a series taken from mean or median or mode. Mean deviation is one of the calculated measures in which all the values are considered in their calculations. It has a series significance as it is an arithmetic mean of the variations of the value of individual items in the series from their central tendency.

Merits

1. It is easy to understand and calculate.
2. Mean deviation is less affected by the extreme values as compared to standard deviations.
3. Mean deviation about an arbitrary point is least when the point is median.

Demerits

1. In mean deviation the signs of all deviations are taken as positive and therefore it is not suitable for further algebraic treatment.
2. This method may not yield accurate results.
3. It yields best results when taken from the median, but the median itself is not a satisfactory average where variability is too high.

Uses

1. Mean deviation and its coefficient are used in studying economic problems such as distribution of income and wealth in a society.

Individual series

Signs (plus and minus) of deviations are disregarded and absolute values of the deviations are summed up. Symbolically, we use $|x_i - \bar{x}|$, which means the deviation of the ith observation of x from the central value $\bar{x}$, (which may be mean or median or mode) with positive sign. Here the vertical lines stand for positive value. Now add up all n observations to get $\sum_{i=1}^{n} |x_i - \bar{x}|$.

Then the mean deviation about mean (MD $\bar{x}$) = $\dfrac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$

Here n is the number of observations.

EXAMPLE 6.7: Find the mean deviation about mean of the marks obtained by 10 students of 10th class in Mathematics in an examination. The marks obtained are 25, 30, 21, 55, 47, 10, 15, 17, 45, 35.

Solution: Let the marks be denoted by x.
Given that x values are 25, 30, 21, 55, 47, 10, 15, 17, 45, 55.

$$\sum_{i=1}^{n} x_i = 25 + 30 + 21 + 55 + 47 + 10 + 15 + 17 + 45 + 55 = 300$$

$$n = 10$$

Arithmetic mean ($\bar{x}$) $\dfrac{300}{10} = 30$

Now

x	$x - \bar{x}$	$\lvert x - \bar{x} \rvert$
25	$25 - 30 = -5$	5
30	$30 - 30 = 0$	0
21	$21 - 30 = -9$	9
55	$55 - 30 = 25$	25
47	$47 - 30 = 17$	17
10	$10 - 30 = -20$	20
15	$15 - 30 = -15$	15
17	$17 - 30 = -13$	13
45	$45 - 30 = 15$	15
55	$35 - 30 = \cancel{25}\ 5$	25
		144

$$(\text{MD}\ \bar{x}) = \frac{\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert}{n} = \; = \frac{144}{10} = 14.4$$

$$\text{Coefficient of mean deviation} = \frac{\text{Mean deviation}}{\text{Mean}} = \frac{14.4}{30} = 0.48$$

When the number of terms is odd

The terms are arranged in ascending order or descending order and then $\left(\frac{n+1}{2}\right)^{th}$ the term value is taken of the order as median.

EXAMPLE 6.8: Find mean deviation about median from the following data.

17	19	21	13	16	18	24	22	20

Solution:

Given values are 17, 19, 21, 13, 16, 18, 24, 22, 20

First arrange the terms in ascending order or descending order

13	16	17	18	19	20	21	22	24

Total number of values $n = 9$ (odd number)

$$\text{Median} = \left(\frac{n+1}{2}\right) = \left(\frac{9+1}{2}\right) = \frac{10}{2} \text{ 5th value of the order is 19}$$

$\therefore$ Median = 19

Median is denoted by M_d

Now

x	$x - M_d$	$\lvert x - M_d \rvert$
17	$17 - 19 = -2$	2
19	$19 = 19 = 0$	0
21	$21 - 19 = 2$	2
13	$13 - 19 = -6$	6
16	$16 - 19 = -3$	3
18	$18 - 19 = -1$	1
24	$24 - 19 = 5$	5
22	$22 - 19 = 3$	3
20	$20 - 19 = 1$	1
		23

$$\text{MD about median} = \frac{\sum_{i=1}^{n} \lvert x_i - M_d \rvert}{n} = \frac{23}{9} = 2.5556$$

$$\text{Coefficient of mean deviation} = \frac{\text{Mean deviation}}{\text{Median}} = \frac{2.5556}{19} = 0.1345$$

When the number of terms is even

The terms are arranged in ascending order or descending order and then the mean of two middle terms is taken as median

i.e. median $= \dfrac{\left(\frac{n}{2}\right)+\left(\frac{n}{2}+1\right)}{2}$ value of the order

where n is number of items.

EXAMPLE 6.9: Find the median of the following data.

87	71	32	28	69	85	53	90	60	56

Solution:

Given values are: 87, 71, 32, 28, 69, 85, 53, 90, 60, 56.
First arrange the terms in ascending or descending order

28	32	53	56	60	69	71	85	87	90

Total number of values $n = 10$ (even number)

$$\left(\frac{n}{2}\right) = \left(\frac{10}{2}\right) \text{ 5th value of the order is 60}$$

$$\left(\frac{n}{2}+1\right) = \left(\frac{10}{2}+1\right) = (5 + 1)\text{th value of the order is 69}$$

i.e. median = $\dfrac{\left(\frac{n}{2}\right)+\left(\frac{n}{2}+1\right)}{2}$ value of the order

$$\text{Median} = \left(\frac{60+69}{2}\right) = \frac{129}{2} = 64.5$$

Median = 64.5

Median is denoted by M_d

x	$x - M_d$	$\lvert x - M_d \rvert$
87	$87 - 64.5 = 22.5$	22.5
71	$71 - 64.5 = 6.5$	6.5
32	$32 - 64.5 = -32.5$	32.5
28	$28 - 64.5 = -36.5$	36.5
69	$69 - 64.5 = 4.5$	4.5
85	$85 - 64.5 = 20.5$	20.5
53	$53 - 64.5 = -11.5$	11.5
90	$90 - 64.5 = 25.5$	25.5
60	$60 - 64.5 = -4.5$	4.5
56	$56 - 64.5 = -8.5$	8.5
		173

$$\text{MD about median} = \frac{\sum_{i=1}^{n} |x_i - M_d|}{n} = \frac{173}{10} = 17.3$$

$$\text{Coefficient of mean deviation} = \frac{\text{Mean deviation}}{\text{Median}} = \frac{1.73}{64.5} = 0.02682$$

EXAMPLE 6.10: Find the mean deviation about mode from the following data.

2	4	5	7	7	7	9	6	8

Solution: In this data 7 is repeated 3 times

So mode is 7.

Mode is denoted by M_o

Now

x	$x - M_o$	$\|x - M_o\|$
2	$2 - 7 = -5$	5
4	$4 - 7 = -3$	3
5	$5 - 7 = -2$	2

Contd.

7	7 – 7 = 0	0
7	7 – 7 = 0	0
7	7 – 7 = 0	0
9	9 – 7 = 2	2
6	6 – 7 = – 1	1
8	8 – 7 = 2	2
		15

$$\text{MD about mode} = \frac{\sum_{i=1}^{n} |x_i - M_d|}{n} = \frac{15}{9} = 1.6667$$

$$\text{Coefficient of mean deviation} = \frac{\text{Mean deviation}}{\text{Mode}} = \frac{1.6667}{7} = 0.2381$$

Discrete series

EXAMPLE 6.11: Find the mean deviation about arithmetic mean from the frequency tabl

Marks	30	40	50	60	70	80	90
No. of students	15	20	10	15	20	15	5

Solution: Let the marks be denoted by x and the number of students by frequencies f.

x	f	fx
30	15	$30 \times 15 = 450$
40	20	$40 \times 20 = 800$
50	10	$50 \times 10 = 500$
60	15	$60 \times 15 = 900$
70	20	$70 \times 20 = 1400$
80	15	$80 \times 15 = 1500$
90	5	$90 \times 5 = 450$
	100	5700

$$\text{Arithmetic mean } (\overline{x}) = \frac{\sum f_i x_i}{\sum f_i}$$

Here $\Sigma f_i x_i = 5700$ and $\Sigma f_i = N = 100$

$$\text{Arithmetic mean } (\overline{x}) = \frac{5700}{100} = 57$$

Thus, arithmetic mean is 57.

Now

x	f	$\lvert x - \bar{x} \rvert$	$f \lvert x - \bar{x} \rvert$
30	15	$\lvert 30 - 57 \rvert = 27$	$15 \times 27 = 405$
40	20	$\lvert 40 - 57 \rvert = 17$	$20 \times 17 = 340$
50	10	$\lvert 50 - 57 \rvert = 7$	$10 \times 7 = 70$
60	15	$\lvert 60 - 57 \rvert = 3$	$15 \times 3 = 45$
70	20	$\lvert 70 - 57 \rvert = 13$	$20 \times 13 = 260$
80	15	$\lvert 80 - 57 \rvert = 23$	$15 \times 23 = 345$
90	5	$\lvert 90 - 57 \rvert = 33$	$5 \times 33 = 165$
	100		1640

$$\text{Mean deviation about mean} = \frac{\sum f \lvert x - \bar{x} \rvert}{N} = \frac{1640}{100} = 16.40$$

Here $$N = \Sigma f_i$$

$$\text{Coefficient of mean deviation} = \frac{\text{Mean deviation}}{\text{Mean}} = \frac{16.40}{57} = 0.2877$$

EXAMPLE 6.12: Calculate mean deviation about median for the following data.

No. of students	6	4	16	7	8	2
Marks	20	9	25	50	40	80

Solution: Let the number of students be denoted by f, marks by x, and cumulative frequency by cf.

Arrange the marks in ascending order and preparing the table as follows.

x	f	cf
9	4	$0 + 4 = 4$
20	6	$4 + 6 = 10$
25	16	$10 + 16 = 26$
40	8	$26 + 8 = 34$
50	7	$34 + 7 = 41$
80	2	$41 + 2 = 43$
	43	

Total frequency = $N = 43$

$$\text{Median} = \text{Size of}\left(\frac{N+1}{2}\right)^{\text{th}} \text{item} = \text{Size of}\left(\frac{43+1}{2}\right) = \left(\frac{44}{2}\right) = 22.$$

The above table shows that all items from 11 to 26 have their values of the variable x as 25. Since 25 items lies in this interval, therefore, its value is 25

Hence, median = 25 marks

Median is denoted by M_d

x	f	$\lvert x - M_d \rvert$	$f\lvert x - M_d \rvert$
9	4	$\lvert 9 - 25 \rvert = 16$	$4 \times 16 = 64$
20	6	$\lvert 20 - 25 \rvert = 5$	$6 \times 5 = 30$
25	16	$\lvert 25 - 25 \rvert = 0$	$16 \times 0 = 0$
40	8	$\lvert 40 - 25 \rvert = 15$	$8 \times 15 = 120$
50	7	$\lvert 50 - 25 \rvert = 25$	$7 \times 25 = 175$
80	2	$\lvert 80 - 25 \rvert = 55$	$2 \times 55 = 110$
	43		499

$$\text{Mean deviation about median} = \frac{\sum f\lvert x - M_d \rvert}{N} = \frac{499}{43} = 11.6047$$

$$\text{Coefficient of mean deviation} = \frac{\text{Mean deviation}}{\text{Median}} = \frac{11.6047}{25} = 0.4642$$

EXAMPLE 6.13: From the following data find the mean deviation about mode of the ages of different persons.

Age	15	20	25	30	35	40	45	50	55
No. of persons	2	3	4	10	11	12	3	2	1

Solution: In this case, it is not possible to state by inspection whether the mode is 30, 35 or 40. We need to determine it through grouping and analysis tables.

Grouping Table: Frequency

Age (x)	*Frequency (f) – I*	*Of two II*	*Of two leaving the first III*	*Of three IV*	*Of three leaving the first V*	*Of three leaving the first two VI*
15	2					
20	3	5		9		
25	4		7			
30	10	14		33	17	25
35	11		21			
40	12	23		6		
45	3		15		26	
50	2	5				17
55	1		3			

Analysis Table: Age in years

Column No.	*I*	*II*	*III*	*IV*	*V*	*VI*	*Total*
15							–
20							–
25						x	1
30			x	x		x	3
35		x	x	x	x	x	5
40	x	x		x	x		4
45					x		1
50							–
55							–

Since the value 35 has occurred the maximum number of times, i.e. 5, the modal age is 35.
Mode is (M_o) = '35'

x	f	$\lvert x - M_o \rvert$	$f\lvert x - M_o \rvert$
15	2	$\lvert 15 - 35 \rvert = 20$	$2 \times 20 = 40$
20	3	$\lvert 20 - 35 \rvert = 15$	$3 \times 15 = 45$
25	4	$\lvert 25 - 35 \rvert = 10$	$4 \times 10 = 40$
30	10	$\lvert 30 - 35 \rvert = 5$	$10 \times 5 = 50$
35	11	$\lvert 35 - 35 \rvert = 0$	$11 \times 0 = 0$
40	12	$\lvert 40 - 35 \rvert = 5$	$12 \times 5 = 60$
45	3	$\lvert 45 - 35 \rvert = 10$	$3 \times 10 = 30$
50	2	$\lvert 50 - 35 \rvert = 15$	$2 \times 15 = 30$
55	1	$\lvert 55 - 35 \rvert = 20$	$1 \times 20 = 20$
	48		315

$$\text{Mean deviation about mode} = \frac{\sum f\lvert x - M_d \rvert}{N} = \frac{315}{48} = 6.5625$$

$$\text{Coefficient of mean deviation} = \frac{\text{Mean deviation}}{\text{Mode}} = \frac{6.5625}{35} = 0.1875$$

Continuous series

EXAMPLE 6.14: Calculate mean deviation about mean for the following data.

Class intervals	0–10	10–20	20–30	30–40	40–50	50–60
Frequency	4	8	14	10	9	5

Solution: Let the class interval be denoted by CI and frequency by f.
x is denoted by mid values of the class intervals.

CI	x	f	$f \times x$
0–10	$\frac{0+10}{2} = \frac{10}{2} = 5$	4	$4 \times 5 = 20$
10–20	$\frac{10+20}{2} = \frac{30}{2} = 15$	8	$8 \times 15 = 120$
20–30	$\frac{20+30}{2} = \frac{50}{2} = 25$	14	$14 \times 25 = 350$
30–40	$\frac{30+40}{2} = \frac{70}{2} = 35$	10	$10 \times 35 = 350$
40–50	$\frac{40+50}{2} = \frac{90}{2} = 45$	9	$9 \times 45 = 405$
50–60	$\frac{50+60}{2} = \frac{110}{2} = 55$	5	$5 \times 55 = 275$
		50	$= 1520$

Here $N = 50, \Sigma fx = 1520$

$$\bar{x} = \frac{\sum fx}{N} = \frac{1520}{50} = 30.4$$

Mean $(\bar{x}) = 30.4$

Now

CI	x	f	$\lvert x - \bar{x} \rvert$	$f \lvert x - \bar{x} \rvert$
0–10	5	4	$\lvert 5 - 30.4 \rvert = 25.4$	$4 \times 25.4 = 101.6$
10–20	15	8	$\lvert 15 - 30.4 \rvert = 15.4$	$8 \times 15.4 = 123.2$
20–30	25	14	$\lvert 25 - 30.4 \rvert = 5.4$	$14 \times 5.4 = 75.6$
30–40	35	10	$\lvert 35 - 30.4 \rvert = 4.6$	$10 \times 4.6 = 46.0$
40–50	45	9	$\lvert 45 - 30.4 \rvert = 14.6$	$9 \times 14.6 = 131.4$
50–60	55	5	$\lvert 55 - 30.4 \rvert = 24.6$	$5 \times 24.6 = 123.0$
		50		600.8

$$\text{Mean deviation about mean} = \frac{\sum f \lvert x - \bar{x} \rvert}{N} = \frac{600.8}{50} = 12.016$$

$$\text{Coefficient of mean deviation} = \frac{\text{Mean deviation}}{\text{Mean}} = \frac{12.016}{30.4} = 0.3953$$

EXAMPLE 6.15: The following table gives the marks obtained by students in Chemistry. Find the median.

Marks	10–19	20–29	30–39	40–49	50–59	60–69	70–79
No. of students	7	16	24	27	22	15	9

Solution: Let the marks be denoted by class intervals (CI) and the number of students by frequency f.

CI	*Class boundary*	f	cf
10–19	9.5–19.5	7	$0 + 7 = 7$
20–29	19.5–29.5	16	$7 + 16 = 23$
30–39	29.5–39.5	24	$23 + 24 = 47$
40–49	39.5–49.5	27	$47 + 27 = 74$
50–59	49.5–59.5	22	$74 + 22 = 96$
60–69	59.5–69.5	15	$96 + 15 = 111$
70–79	69.5–79.5	9	$111 + 9 = 120$
		120	

Here $N = 120, \dfrac{N}{2} = \dfrac{120}{2} = 60, m = 47, f = 27, l = \dfrac{40 + 39}{2} = \dfrac{79}{2} = 39.5, C = 10$

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times C = 39.5 + \frac{60 - 47}{27} \times 10 = 39.5 + \frac{13}{27} \times 10 = 39.5 + 4.8148$$

Median $= 44.3148$

Now

CI	x	f	$\lvert x - M_d \rvert$	$f \lvert x - M_d \rvert$
10–19	14.5	7	$\lvert 14.5 - 44.32 \rvert = 29.82$	$7 \times 29.82 = 208.74$
20–29	24.5	16	$\lvert 24.5 - 44.32 \rvert = 19.82$	$16 \times 19.82 = 317.12$
30–39	34.5	24	$\lvert 34.5 - 44.32 \rvert = 9.82$	$24 \times 9.82 = 235.68$
40–49	44.5	27	$\lvert 44.5 - 44.32 \rvert = 0.18$	$27 \times 0.18 = 4.86$
50–59	54.5	22	$\lvert 54.5 - 44.32 \rvert = 10.18$	$22 \times 10.18 = 223.96$
60–69	64.5	15	$\lvert 64.5 - 44.32 \rvert = 20.18$	$15 \times 20.18 = 302.70$
70–79	74.5	9	$\lvert 74.5 - 44.32 \rvert = 30.18$	$9 \times 30.18 = 271.62$
		120		1564.68

$$\text{Mean deviation about median} = \frac{\sum f \lvert x - M_d \rvert}{N} = \frac{1564.68}{120} = 13.039$$

$$\text{Coefficient of mean deviation} = \frac{\text{Mean deviation}}{\text{Median}} = \frac{13.039}{44.32} = 0.2942$$

EXAMPLE 6.16: From the following table, calculate the mean deviation about mode.

Class intervals	0–10	10–20	20–30	30–40	40–50	50–60	60–70
Frequency	9	11	14	20	15	10	8

Solution:

Class intervals	0–10	10–20	20–30	30–40	40–50	50–60	60–70
Frequency	9	11	14	20	15	10	8

l = Lower limit of the modal class $= \frac{30+30}{2} = \frac{60}{2} = 30$

f = Maximum frequency = 20

f_1 = Maximum frequency of the class preceding value = 14

f_2 = Maximum frequency of the class succeeding value = 15

C = Class interval size = 10

$$\text{Mode} = l + \frac{f - f_1}{2f - f_1 - f_2} \times C = 30 + \frac{20-14}{2 \times 20 - 14 - 15} \times 10 = 30 + \frac{60}{11} \times 10 = 35.4545 = 35.46$$

Mode $(M_o) = 35.46$

Now

CI	x	f	$\lvert x - M_o \rvert$	$f\lvert x - M_o \rvert$
0–10	5	9	$\lvert 5 - 35.46 \rvert = 30.46$	$9 \times 30.46 = 274.14$
10–20	15	11	$\lvert 15 - 35.46 \rvert = 20.46$	$11 \times 20.46 = 225.06$
20–30	25	14	$\lvert 25 - 35.46 \rvert = 10.46$	$14 \times 10.46 = 146.44$
30–40	35	20	$\lvert 35 - 35.46 \rvert = 0.46$	$20 \times 0.46 = 9.20$
40–50	45	15	$\lvert 45 - 35.46 \rvert = 10.46$	$15 \times 10.46 = 156.90$
50–60	55	10	$\lvert 55 - 35.46 \rvert = 20.46$	$10 \times 20.46 = 204.60$
60–70	65	8	$\lvert 65 - 35.46 \rvert = 30.46$	$8 \times 30.46 = 243.68$
		87		1260.02

$$\text{Mean deviation about mode} = \frac{\sum f\lvert x - M_o \rvert}{N} = \frac{1260.02}{87} = 14.4830$$

$$\text{Coefficient of mean deviation} = \frac{\text{Mean deviation}}{\text{Mode}} = \frac{14.4830}{35.46} = 0.4084$$

4. Standard Deviation

The idea of standard deviation was first presented by Karl Pearson in 1893. This measure is widely used for studying dispersion. Standard deviation is the positive square root of the average of squared deviations taken from arithmetic mean. It is generally denoted by the Greek alphabet σ or by SD. Let x be a random variate which takes on n values, namely $x_1, x_2, \ldots, x_n$, then the standard deviation of these n observations is given by

$$\sigma = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}}, \quad \text{where } \bar{x} = \frac{\sum x}{n} \text{ OR}$$

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} \quad \text{OR} \quad \sigma = \sqrt{\frac{\sum x^2}{n} - (\bar{x})^2}$$

Standard deviation is also known as root mean square deviation.

Coefficient of standard deviation $= \frac{\sigma}{\bar{x}}$

Merits

1. It is based on all the observations.
2. It is rigidly defined.
3. It represents the true measurement of dispersion of series.
4. It is least affected by fluctuation of sampling.
5. It has a greater mathematical significance and is capable of further mathematical treatments.

Demerits

1. It is difficult to compute unlike other measures of dispersion.
2. It is not easily understood.
3. It gives more weightage to extreme values.
4. It consumes much time and labour while computing it.

Uses

1. It is widely used in biological studies.
2. It is used in fitting a normal curve to a frequency distribution.
3. It is the most widely used measure of dispersion.

Individual series

EXAMPLE 6.17: Find the standard deviation and its coefficient for the following data.

X: 11, 13, 14, 15, 17, 18, 19, 20.

Solution:

$$\Sigma x = 11 + 13 + 14 + 15 + 17 + 18 + 19 + 20 = 127$$

$$\Sigma x^2 = (11)^2 + (13)^2 + (14)^2 + (15)^2 + (17)^2 + (18)^2 + (19)^2 + (20)^2$$

$$\Sigma x^2 = 2085.$$

$$n = \text{Number of samples} = 8$$

$$\text{Mean } (\bar{x}) = \frac{\sum x}{n} = \frac{127}{8} = 15.875$$

$$\sigma = \sqrt{\frac{\sum x^2}{n} - (\bar{x})^2} = \sqrt{\frac{2085}{8} - (15.875)^2} = \sqrt{260.625 - 252.016} = \sqrt{8.609} = 2.9341$$

Mean $(\bar{x}) = 15.875$ and $SD(\sigma) = 2.9341$

$$\text{Coefficient of standard deviation} = \frac{\sigma}{\bar{x}} = \frac{2.9341}{15.875} = 0.1848$$

Discrete series

EXAMPLE 6.18: Find the standard deviation and its coefficient from the following data.

Size of the item	10	11	12	13	14	15	16
Frequency	2	7	11	15	10	4	1

Solution: Let the size of the item be denoted by x and frequency by f.

x	f	$fx = (f)(x)$	$fx^2 = (f)(x)(x)$
10	2	$2 \times 10 = 20$	$2 \times 10 \times 10 = 200$
11	7	$7 \times 11 = 77$	$7 \times 11 \times 11 = 847$
12	11	$11 \times 12 = 132$	$11 \times 12 \times 12 = 1584$
13	15	$15 \times 13 = 195$	$15 \times 13 \times 13 = 2535$
14	10	$10 \times 14 = 140$	$10 \times 14 \times 14 = 1960$
15	4	$4 \times 15 = 60$	$4 \times 15 \times 15 = 900$
16	1	$1 \times 16 = 16$	$1 \times 16 \times 16 = 256$
	50	640	8282

Here $\Sigma f = N = 50, \Sigma fx = 640, \Sigma fx^2 = 8282$

$$\text{Mean } (\bar{x}) = \frac{\sum fx}{N} = \frac{640}{50} = 12.8$$

$$\text{Standard deviation } (\sigma) = \sqrt{\frac{\sum fx^2}{N} - (\bar{x})^2} = \sqrt{\frac{8282}{50} - (12.8)^2} = \quad 1.4142$$

$$\text{Mean } (\bar{x}) = 12.8 \text{ and SD } (\sigma) = 1.4142$$

$$\text{Coefficient of standard deviation} = \frac{\sigma}{\bar{x}} = \frac{1.4142}{12.8} = \quad 0.1105$$

Continuous series

EXAMPLE 6.19: Calculate standard deviation and its coefficient for the following data.

Profits (Rs.)	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100
No. of companies	30	58	62	85	112	70	57	26

Let the profits be denoted by CI and the number of companies by f.
x mid values of the class intervals

CI	x	f	$d = \frac{x-A}{c}$	fd	fd^2
20–30	25	30	$\frac{25-65}{10} = -4$	–120	480
30–40	35	58	$\frac{35-65}{10} = -3$	–174	522
40–50	45	62	$\frac{45-65}{10} = -2$	–124	248
50–60	55	85	$\frac{55-65}{10} = -1$	–85	85
60–70	65	112	$\frac{65-65}{10} = 0$	0	0
70–80	75	70	$\frac{75-65}{10} = 1$	70	70
80–90	85	57	$\frac{85-65}{10} = 2$	114	228
90–100	95	26	$\frac{95-65}{10} = 3$	78	234
				–241	1867

Here $A = 65,\ C = 10,$

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

$$\sigma = 10 \times \sqrt{\frac{1867}{500} - \left(\frac{-241}{500}\right)^2} = 10 \times \sqrt{3.734 - 0.232} = 18.71$$

Mean $(\bar{x}) = A + \frac{\sum fd}{N} \times C = 65 + \left(\frac{-241}{500}\right) \times 10 = 65 - 4.82 = 60.18$

Mean $(\bar{x}) = 60.18$ and $SD(\sigma) = 18.71$

Coefficient of standard deviation $= \frac{\sigma}{\bar{x}} = \frac{18.71}{60.18} = 0.3109$

ariance and coefficient of variation

riance: Variance is the square of standard deviation and is denoted by σ^2. The methods for lculating variance are the same as for the standard deviation.

Coefficient of variation: It is a relative measure of dispersion. It is denoted by CV.

$$\text{Coefficient of variation (CV)} = \frac{\sigma}{\overline{x}} \times 100$$

where σ is the standard deviation and $\overline{x}$ is the mean of the given series. It is important to note that the coefficient of variation is always a percentage.

EXAMPLE 6.20: The following are the runs scored by two batsmen A and B in ten innings.

A	101	27	0	36	82	45	7	13	65	14
B	97	12	40	96	13	8	85	8	56	15

Who is more consistent?

Solution:

Here $\Sigma A = 390$ and $n_1 = 10$, $\overline{x}_A = \dfrac{\sum A}{n_1} = \dfrac{390}{10} = 39$

$\Sigma B = 430$ and $n_2 = 10$, $\overline{x}_B = \dfrac{\sum B}{n_2} = \dfrac{430}{10} = 43$

Batsman A			Batsman B		
Runs scored	$d_x = x - 39$	dx^2	Runs scored	$dy = y - 43$	dy^2
101	62	3844	97	54	2916
27	–12	144	12	–31	961
0	–39	1521	40	–3	9
36	–3	9	96	53	2809
82	43	1849	13	–30	900
45	6	36	8	–35	1225
7	–32	1024	85	42	1764
13	–26	676	8	–35	125
65	26	676	56	13	169
14	–25	625	15	–8	784
	$\Sigma dx = 0$	$\Sigma dx^2 = 10404$		$\Sigma dy = 0$	$\Sigma dy^2 = 12762$

Here $\Sigma dx = 0, \Sigma dx^2 = 10404, \Sigma dy = 0, \Sigma dy^2 = 12762$

$$\sigma_A = \sqrt{\frac{\sum dx^2}{n_1} - \left(\frac{\sum dx}{n_1}\right)^2} = \sqrt{\frac{10404}{10}} = 32.26$$

$$\sigma_B = \sqrt{\frac{\sum dy^2}{n_2} - \left(\frac{\sum dy}{n_2}\right)} = \sqrt{\frac{12762}{10}} = 35.72$$

$$\text{Coefficient of variation of batsman } A = \frac{\sigma_A}{\bar{x}_A} \times 100 = \frac{32.26}{39} \times 100 = 82.72\%$$

$$\text{Coefficient of variation of batsman } B = \frac{\sigma_B}{\bar{x}_B} \times 100 = \frac{35.72}{43} \times 100 = 83.07\%$$

Now, coefficient of variation of batsman of A is less than the coefficient of variation of batsman B. Hence, batsman A is more consistent than batsman B.

END