$$P_{90} = l + \frac{h}{f}\left(\frac{90N}{100} - c.f.\right)$$

$$= 50 + \frac{10}{4}\left(\frac{90 \times 50}{10} - 43\right) = 50 + 5 = 55$$

Thus the ninetieth percentile is 55.

For P_{23} : $\frac{25N}{100} = 12.5$.

Thus the ninetieth percentile class is 10 - 20

$$P_{25} = l + \frac{h}{f}\left(\frac{25N}{100} - c.f.\right)$$

$$= 10 + \frac{10}{10}\left(\frac{25 \times 50}{10} - 3\right) = 10 + 9.5 = 19.5$$

Thus the ninetieth percentile is 19.5

1.28 Mode

According to Coxton and Cowden, "the mode of a distribution is the value at t around which, the items tend to be most heavily concentrated. It may be regarded as t typical of a series of values". The variate value 'x' having maximum frequency in a dis is known as its mode. In other words, the mode is that value of a series which appe frequently than any other. Mode is not necessarily unique, there can be more than one m distribution.

1.28.1 Mode from Grouped Data

(a) **Discrete Series:** Mode of a discrete frequency distribution can be determ inspection. The value having maximum frequency is the mode of the distributio

(b) **Continuous Series:** Mode of a continuous frequency distribution with c ascending order can be calculated by using the following formula:

$$\text{Mode } (M_o) = l + \frac{f_0 - f_1}{2f_0 - f_1 - f_2} \times h$$

where

l = lower limit of the modal class

f_0 = frequency of the modal class

f_1 = frequency of the preceding class to the modal class

f_2 = frequency of the succeeding class to the modal class

h = size of class interval

Remarks

1. While calculating value of mode it should be seen that the class intervals of the different classes are equal otherwise the above formula will not give the correct answer. In this case mode can be find by using an empirical relationship between mean, median and mode as

   Mean - Mode = 3 (Mean - Median)

   or Mode = 3 Median – 2 Mean

2. A frequency distribution having more than one mode is called **multimodal distribution.** If there is only one mode it is called **unimodal distribution and if there are two mode, it is called as bimodal.**

3. The class having the maximum frequency is called the modal class.

1.28.2 Graphical Method for finding the Mode

The steps involved are

(i) Prepare a histogram of the given data.

(ii) The highest rectangle will be the modal class.

(iii) Draw two lines diagonally in the inside of the modal class rectangle to the upper corner of the adjacent bar.

(iv) From the point of intersection of these lines, draw a perpendicular on X-axis which gives the modal value.

1.28.3 Grouping Method for finding the Mode

Grouping method is used when there exists one or more of the following cases:

(i) If the maximum frequency is repeated.

(ii) If the maximum frequency occurs in the very beginning or at the end of the distribution.

(iii) If there are irregularities in the distribution.

In grouping method, the values are first arranged in ascending order of their magnitude, then grouping table is prepared. Normally the grouping table has six columns.

Column 1: Write the original given **frequencies.**

Column 2: Sum of the frequencies taken two at a time starting from first frequency of column I.

Column 3: Leave the first frequency and remaining frequencies are grouped in twos.

Column 4: The frequencies are grouped in threes.

Column 5: Leave the first frequency and remaining frequencies are grouped in threes.

Column 6: Leaving the first two frequencies, the remaining frequencies are grouped in three

Note: **Underline** the highest value of each column.

Now an analysis table is formed using grouping table for finding the value which is repe the maximum number of times. This value is the mode.

1.28.4 Advantages and Disadvantages of Mode

Advantages

1. It is an average of position.
2. It is one of the value in the series.
3. It is not affected by open end classes.
4. It is not at all affected by the extreme values.
5. It can be found by mere inspection in some cases.
6. It is of great importance in manufacturing of garments, shoes etc. due to most comm size.

Disadvantages

1. It is not well defined and is not unique.
2. It is not based on all observations.
3. It is based only on concentrated values.
4. In a number of items, the mode does not exit.
5. It is not capable of further mathematical treatment.
6. It is not rigidly defined. Different methods can give different values of mode.
7. Calculation of mode in case of discrete case is more difficult than the calculation mean.

Illustrative Examples

Example 1. Find the mode from the following data:

11, 28, 11, 23, 35, 35, 11, 17, 23, 12, 6, 11.

Solution Since the observation 11 occurs a maximum number of times.

Thus mode is 11.

Example 2. Find the mode from the following data: 18, 12, 13, 15, 15,16, 13, 12, 16, 12

Solution Since the observation 12 occurs a maximum number of times.

Thus mode is 12.

Example 3. Find the mode from the following frequency distribution:

x	18	20	22	24	26	28	30
f	3	5	9	10	20	9	4

Solution In the given problem mode can be find by inspection.

The variate value 26 has the maximum frequency 20.

Thus Mode = 26.

Example 4. Find the mode from the following data:

28, 16, 18, 13, 15, 16, 26, 15, 16, 18.

Solution Arranging the data as frequency distribution, we have

Value	Frequency
13	1
15	2
16	3
18	2
26	1
28	1

It is clear from the table that the highest frequency is for the value 26.

Thus Mode = 26.

Example 5. Find mode from the following frequency distribution:

Class	52-55	55-58	58-61	61-64
Frequency	15	20	25	10

Solution

Class	Frequency
52-55	15
55-58	$20(f_1)$
58-61	$25(f_0)$
61-64	$10(f_2)$

Here, maximum frequency 25 occurs in the class 58-61, so it is the modal class.

$l = 58, f_0 = 25, f_1 = 20, f_2 = 10$ and $h = 3$

$$\text{Mode } (M_o) = l + \frac{f_0 - f_1}{2f_0 - f_1 - f_2} \times h$$

$$= 58 + \frac{25 - 20}{(2 \times 25) - 20 - 10} \times 3$$

$$= 58 + 0.75 = 58.75$$

Example 6. Find mode from the following frequency distribution:

Class interval	0-10	10-20	20-30	30-40	40-50	50-60
Frequency	8	7	16	24	15	7

Solution

Class Interval	Frequency
0-10	8
10-20	7
20-30	16(f_1)
30-40	24(f_0)
40-50	15(f_2)
50-60	7

Here, maximum frequency 24 occurs in the class 30-40, so it is the modal class.

We have

$l = 30, f_0 = 24, f_1 = 16, f_2 = 15$ and $h = 10$

$$\text{Mode } (M_o) = l + \frac{f_0 - f_1}{2f_0 - f_1 - f_2} \times h$$

$$= 30 + \frac{24 - 16}{(2 \times 24) - 16 - 15} \times 10$$

$$= 34.706$$

Example 7. The expenditure of 100 families is given below:

Expenditure	0-10	10-20	20-30	30-40	40-50
No. of families	14	?	27	?	15

Given that mode is 24. Determine the missing frequencies.

Solution Let the missing frequencies be a and b.

Expenditure	No. of Families
0-10	14
10-20	a (f_1)
20-30	27(f_0)
30-40	b (f_2)
40-50	15

Here, mode is 24, so modal class is 20-30.

$l = 20, f_0 = 27, f_1 = a, f_2 = b$ and $h = 10$

$$\text{Mode } (M_o) = l + \frac{f_0 - f_1}{2f_0 - f_1 - f_2} \times h$$

$$24 = 20 + \frac{27 - a}{(2 \times 27) - a - b} \times 10$$

$$4 = \frac{27-a}{(2\times 27)-a-b}\times 10$$

$$3a-2b=27 \qquad \ldots(i)$$

Since N = 100, we have $56+a+b=100$

$$a+b=44 \qquad \ldots(ii)$$

On solving equations (i) and (ii), we get $a=23$ and $b=21$.

Hence, missing frequencies are 23 and 21.

Example 8. Obtain the mode for the following date

Weekly Wages	10 – 15	15 – 20	20 – 25	25 –30	30 – 40	40 – 60	60 – 80
No. of Workers	7	19	27	15	12	12	8

Solution Rearrange the frequency distribution

Weekly Wages (in ₹)	No of Workers
0 – 20	26
20 – 40	54
40 – 60	12
60 – 80	8

Here $l=20, f_1=26, f_0=54, f_2=12$ and $h=20$

$$\text{Mode } (M_o) = l + \frac{f_0-f_1}{2f_0-f_1-f_2}\times h$$

$$= 20 + \frac{54-26}{(2\times 54)-26-12}\times 20$$

$$= 20+8=28$$

Example 9. Obtain the mode for the following data

Wages (₹) less than	50	100	150	200	250	300	350	400	450
No. of workers	29	224	465	582	634	644	650	653	655

Solution Here the data is given in the form of cumulative frequency. First of all we have to convert it into a normal continuous frequency distribution as follows:

Wages (₹)	No. of Workers (f)
0-50	29
50-100	224-29 = 195(f_1)
100-150	465-224 = 241(f_0)
150-200	582-465 = 117(f_2)
200-250	634-582 = 52
250-300	644-634 = 10
300-350	650-644 = 6
350-400	653-650 = 3
400-450	655-653 = 2

Here maximum frequency 241 occurs in the class 100-150 which is the modal class.

$l = 100, f_0 = 241, f_1 = 195, f_2 = 117$ and $h = 50$

$$\text{Mode } (M_o) = l + \frac{f_0 - f_1}{2f_0 - f_1 - f_2} \times h \qquad = 100 + \frac{241 - 195}{(2 \times 241) - 195 - 117} \times 50$$

Mode =113.5

Example 10. Find the mode of the following frequency distribution by grouping method:

Wages (₹)	125	175	225	275	325	375
No. of Persons	3	8	21	6	4	2

Solution Here the given frequency distribution is not regular since the frequencies are increasing steadily upto 40 and then decreasing upto 20 but the frequency 45 after 20 does not seem to be consistent with the distribution. In this case we use the grouping method.

Grouping Table

Wages	Frequency					
	(1)	(2)	(3)	(4)	(5)	(6)
125	3					
175	8	11				
225	21		29	32		
275	6	27			35	
325	4		10	12		31
375	2	6				

To find the mode we use the following analysis table:

Analysis Table

Column No. (a)	Maximum Frequency (b)	Value or combination of values of x giving maximum frequency in (b) (c)
1	21	**225**
2	27	**225**, 275
3	29	175, **225**, 275
4	32	125, 175, **225**
5	35	175, **225**, 275
6	31	**225**, 275, 325

On examining the values in column (c) of analysis table we find that the value 225 is repeated the maximum number of times. Hence, mode is 225.

Example 11. Find the mode of the following frequency distribution:

x	11	12	13	14	15	16	17	18	19	20	21	22
f	3	8	15	23	35	40	32	28	20	45	14	6

Solution Here the given frequency distribution is not regular since the frequencies are increasing steadily upto 40 and then decreasing upto 20.

But the frequency 45 after 20 does not seem to be consistent with the distribution. In this case we use the grouping method.

Grouping Table

Size (x)	Frequency					
	(1)	(2)	(3)	(4)	(5)	(6)
11	3					
12	8	11		26		
13	15		23		46	
14	23	38	58			73
15	35			98		
16	40	75				
17	32		72		107	
18	28	60		80		100
19	20		48			
20	45				93	
21		65	59			79
22	14			65		
	6	20				

To find the mode we use the following analysis table:

Analysis Table

Column No. (a)	Maximum Frequency (b)	Combination of values of x giving maximum frequency in (b) (c)
1	45	20
2	75	15,16
3	72	16,17
4	98	14,15,16
5	107	15,16,17
6	100	16,17,18

On examining the values in column (c) of analysis table we find that the value is repeated the maximum number of times.

Hence, the value of the mode is 16.

Example 12. Find the mode of the following frequency distribution:

Class interval	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-8
Frequency	9	12	15	16	17	15	10	13

Solution Here the given frequency distribution is not regular we use the grouping method

Grouping Table

Size (x)	Frequency					
	(1)	(2)	(3)	(4)	(5)	(6)
0-10	9					
10-20	12	21				
20-30	15		27	36		
30-40	16	31			43	
40-50	17					48
50-60	15		33	48		
60-70	10	32			42	
70-80	13	23	25			38

To find the mode we use the following analysis table:

Analysis Table

Column No. (a)	Maximum Frequency (b)	Size of the group of containing maximum frequency in (b) (c)
1	17	40-50
2	32	40-50 50-60
3	33	30-40 40-50
4	48	30-40 40-50 50-60
5	43	10-20 20-30 30-40
6	48	20-30 30-40 40-50
No. of times size groups occurs		1 2 4 5 2

Here, the size group 40-50 is occurring maximum number of times (i.e., 5 times) and therefore 40-50 is the modal class.

Here, $l = 40, f_0 = 17, f_1 = 16, f_2 = 15$ and $h = 10$

$$\text{Mode } (M_o) = l + \frac{f_0 - f_1}{2f_0 - f_1 - f_2} \times h$$

$$= 40 + \frac{17-16}{(2\times17)-16-15} \times 10$$

Mode = 43.33

Example 13. Find out the mean, median and mode from the following frequency distribution:

Pension (in ₹)	No of Persons
Less than 10	15
Less than 20	35
Less than 30	60
Less than 40	84
Less than 50	96
Less than 60	127
Less than 70	198
Less than 80	250

Solution

Pension (x)	c.f.	f	$u = \frac{x-35}{10}$	fu
0 – 10	15	15	-3	-45
10 – 20	35	20	-2	-40
20 – 30	60	25	-1	-25
30 – 40	84	24	0	0
40 – 50	96	12	1	12
50 – 60	127	31	2	62
60 – 70	198	71	3	213
70 – 80	250	52	4	208
		$N = 250$		$\sum fu = 385$

Mean A.M. $(\bar{x}) = a + \frac{h}{N}\sum_{i=1}^{8} f_i u_i$

$$(\bar{x}) = 35 + \frac{10}{250}(385) = 50.40$$

Mean $= 50.40$

So Mean Pension is ₹ 50.40

Median here $\frac{N}{2} = 125$, so the median class is 50 – 60

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

$$= 50 + \frac{10}{31}(125 - 96) = 59.35$$

Median $= 59.35$

So Median Pension is ₹ 59.35

Mode $= 3\text{Median} - 2\text{Mean}$

$= 3(59.35) + 2(50.40) = 178.05 + 100.80$

$= 278.85$

So Modal Pension is ₹ 59.35

Example 14. Find out the mode from the following frequency distribution:

Central Sizes	1	2	3	4	5	6	7	8	9	10
Frequency	8	6	10	12	20	12	5	3	2	4

Solution Here the given frequency distribution is not regular we use the grouping method.

Grouping Table

Class Intervals	Frequency					
	(1)	(2)	(3)	(4)	(5)	(6)
0.5 – 1.5	8					
1.5 – 2.5	6	14		24		
2.5 – 3.5			16		28	
3.5 – 4.5	10	22				42
4.5 – 5.5	12					
5.5 – 6.5	20		32	44	37	
6.5 – 7.5	12	32				
7.5 – 8.5	5		17		9	20
8.5 – 9.5	3			10		
9.5–10.5		8	5			
	2					
	4	6				

To find the mode we use the following analysis table:

Column No.	2.5-3.5	3.5-4.5	4.5-5.5	5.5-6.5	6.5-7.5
1			1		
2			1	1	
3		1	1		
4		1	1	1	
5			1	1	1
6	1	1	1		
Total	1	3	6	3	1

Here, by inspection, the mode lies in the class interval; 4.5-5.5 as the total for this class interval is maximum, i.e., 6.

$l = 4.5, f_0 = 12, f_1 = 20, f_2 = 12$ and $h = 1$

$$\text{Mode } (M_o) = l + \frac{f_0 - f_1}{2f_0 - f_1 - f_2} \times h$$

$$= 4.5 + \frac{20 - 12}{(2 \times 20) - 12 - 12} \times 1$$

$$= 4.5 + 0.5 = 5.0$$

$$H.M. = \frac{n}{\sum\left(f \times \frac{1}{x}\right)}$$

$$\Rightarrow H.M. = \frac{88}{4.0943} = 21.493$$

Example 5. Find the harmonic mean for the following distribution.

Marks	0-10	10-20	20-30	30-40	40-50	50 -60
No. of students	15	20	17	13	12	12

Solution

Marks	Mid Points (x)	f	$\frac{1}{x}$	$f \times \frac{1}{x}$
0 – 10	5	15	0.2	3
10 – 20	15	20	0.067	1.34
20 – 30	25	17	0.04	0.68
30 – 40	35	13	0.029	0.377
40 – 50	45	12	0.022	0.264
50 – 60	55	12	0.018	0.216
			$\sum\left(\frac{f}{x}\right) = 5.877$	

$$H.M. = \frac{N}{\sum\left(f \times \frac{1}{x}\right)}$$

$$\Rightarrow H.M. = \frac{89}{5.877} = 15.143$$

1.30 Geometric Mean

Geometric mean is defined as the nth root of the product of n items or values. If there are two items a and b, then we take square root as $\sqrt{ab}$

When there are three items a, b and c, then we take cube root as $\sqrt{abc}$.

Thus if $x_1, x_2, x_3, \ldots\ldots\ldots\ldots, x_n$ are n values of a variate x, none of them being zero, then geometric mean is calculated as

$$G = (x_1 . x_2 . x_3 \ldots\ldots\ldots\ldots\ldots\ldots x_n)^{1/n},$$

Then
$$\log G = \log\left[(x_1 x_2 x_3 \ldots\ldots x_n)^{1/n}\right]$$
$$\log G = \frac{1}{n}\left[\log(x_1 x_2 x_3 \ldots\ldots x_n)\right]$$
$$\log G = \frac{1}{n}(\log x_1 + \log x_2 + \log x_3 + \ldots\ldots + \log x_n)$$
$$G = anti\log\left[\frac{(\log x_1 + \log x_2 + \log x_3 + \ldots\ldots + \log x_n)}{n}\right]$$

1.30.1 Geometric Mean in Case of Frequency Distribution

In case of a frequency distribution, geometric mean of n values $x_1, x_2, x_3 \ldots\ldots$ variate x occurring with frequency $f_1, f_2, f_3 \ldots\ldots, f_n$ respectively is given by:

$$G = anti\log\left[\frac{(f_1 \log x_1 + f_2 \log x_2 + f_3 \log x_3 + \ldots\ldots + f_n \log x_n)}{n}\right]$$

$$G = anti\log\left[\frac{\sum_{i=1}^{n} f_i \log x_i}{n}\right]$$

Thus geometric mean can also be defined as the anti log of weighted mean of the diffe values of $\log x_i$, whose weights are their frequencies f_i. In case of continuous or gro frequency distribution, the values of the variate x are taken to be the values corresponding to mid points of the class intervals. If $\log G_1, \log G_2, \log G_3, \ldots, \log G_n$ are various geometric me of observations $1, 2, 3 \ldots\ldots n$, then combine geometric mean can be obtained as:

$$\log G = \frac{n_1 \log G_1 + n_2 \log G_2 + n_3 \log G_3 + \ldots\ldots + n_n \log G_n}{n_1 + n_2 + \ldots\ldots + n_n}$$

1.30.2 Advantages and Disadvantages of Geometric Mean

Advantages

1. It is based on all the observations.
2. It is useful in the construction of the index numbers.
3. It is not much affected by the fluctuations of sampling.
4. It is very simple and lends s itself to algebraic treatments.
5. It gives less weightage to large numbers and more weightage to small numbers.

Disadvantages

1. It is very difficult to understand
2. It is relatively difficult to compute.
3. It cannot be obtained by inspection.
4. It is impossible to calculate if any of the items is zero or negative.

Illustrative Examples

Example 1. Find out the geometric mean of the following: 1, 2, 3

Solution $G.M. = \sqrt[3]{abc} = \sqrt[3]{1 \times 2 \times 3} = \sqrt[3]{6} = 1.8171$

Example 2. Find out the geometric mean of the following: 2, 4, 8

Solution $G.M. = \sqrt[3]{abc} = \sqrt[3]{2 \times 4 \times 8} = 4$

Example 3. Find out the geometric mean of the following: 12, 13, 14, 15, 16, 17

Solution

x	*logx*
12	1.0792
13	1.1139
14	1.1461
15	1.1761
16	1.2041
17	1.2304
$n = 6$	$\sum \log x = 6.9498$

$$\log G = \frac{1}{n} \sum \log x$$

$$= \frac{1}{6} \times 6.9498 = 1.1583$$

$$\Rightarrow G = anti \log(1.1583) = 14.397$$

Example 4. The annual rates of growth of output of a company in 5 years are: 6, 6.5, 3, 2, 10 percent respectively. What is compound rate of growth of per annum for the period?

Solution Let the starting production of company be 100

Year	*x*	*logx*
Ist Year	106	2.02530
IInd Year	106.5	2.02734
IIIrd Year	103	2.01837
IVth Year	102	2.00860
Vth Year	110	2.04139
		$\sum \log x = 10.121$

$$\log G = \frac{1}{n} \sum \log x$$

$$= \frac{1}{5} \times 10.121 = 2.0242$$

$$\Rightarrow G = anti\log(2.0242) = 105.73$$

Thus percentage increase = 105.73 – 100 = 5.73%

Example 5. Find out the geometric mean of the following:
6, 15, 20, 35, 90, 250, 300, 400, 500.

Solution

x	$\log x$
6	0.7782
15	1.1761
20	1.3010
35	1.5441
90	1.9542
250	2.3979
300	2.4771
400	2.6021
500	2.6987
$n = 6$	$\sum \log x = 16.9294$

$$\log G = \frac{1}{n} \sum \log x$$

$$= \frac{1}{10} \times 16.9294 = 1.69294$$

$$\Rightarrow G = anti\log(1.69294) = 49.31057$$

EXERCISE

1. Describe any two limitations of statistics.
2. Comment on the following
   "Statistics is the science of counting."
3. What is statistics and write its limitations.
4. Comment on the following
   "Statistics is the science of averages."
5. Write the definition of statistics by A.L. Bowley.
6. Comment on the following
   "Statistics is the science of estimates and probabilities."
7. Statistics is all-pervading". Elucidate this statement.

80. Find out the mode from the following frequency distribution:

Marks	No of Students
Below 25	8
Below 30	23
Below 35	51
Below 40	81
Below 45	103
Below 50	113
Below 55	117
Below 60	120

Ans. 36 years

1.31 Measures of Dispersion

According to Spiegel, "the degree to which numerical data tend to spread about an average value is called the variations or dispersion". A measure of dispersion describes the degree of scatter shown by the observations and is usually measured as an average deviation about some central value. A measure of dispersion gives us additional information that enables us to judge the reliability of our measure of central value. It makes possible to compare two series of data in respect of their variability. The main objectives of measuring variability are

1. Control of Variation
2. Comparison of Series
3. Reliability of Average
4. Further Statistical Analysis

1.32 Absolute Measure and Relative Measure

Absolute variability is measured in the same units as the data. For example, if the original data is in kilometer, then the absolute measure is also in kilometer. For this reason, absolute dispersion cannot be used to compare the scatter or variability in series where units are considered different. For comparing two or more series where units are different, relative measure is used as they are calculated as percentage or the coefficient of the absolute measure of dispersion. Therefore, it is called as **coefficient of dispersion**.

1.33 Methods of Measuring Dispersion or Variation

Following are the most commonly used measures of dispersion:

(i) Range
(ii) Interquartile range and quartile deviation
(iii) Mean deviation
(iv) Standard deviation
(v) Coefficient of variation

1.34 Range

Range is the simplest measure of dispersion. It is the difference between the largest (hig and smallest value of the observation. It is actually determined by two extreme values o observations.

Range $= L - S$, where, L = Largest value and S = Smallest value

$$\text{Coefficient of range} = \frac{L-S}{L+S}$$

1.34.1 Advantages and disadvantages of Range

Advantages

1. Range is the simplest measure of dispersion.
2. It is used in quality control for drawing R-charts.
3. It is easy to calculate and gives a broad picture of data very quickly.
4. It is used to study variations in the prices of commodities and movement in the price securities.
5. It is used in weather forecasting e.g., it gives an idea of the variation between maxim and minimum levels of temperature.

Disadvantages

1. It is affected by sampling fluctuation.
2. It is not based on every item of the series.
3. It is not useful for frequency distribution.
4. It is very much affected by the extreme values.
5. It not suitable for further mathematical treatment.
6. It cannot be calculated in case of open end distributions.
7. It depends only on two values (largest and smallest) and ignores all other values highly misleading.

Illustrative Examples

Example 1. Find the Range for the following data:

19, 11, 12, 13, 22, 24, 23, 25, 15, 30, 21

Solution Range $= L - S = 30 - 11 = 19$

Example 2. Find the Range and the coefficient of Range for the following data:

10, 20, 22, 24, 30, 35, 40, 45, 50, 55, 60

Solution Range $= L - S = 60 - 10 = 50$

$$\text{Coefficient of range} = \frac{L-S}{L+S} = \frac{60-10}{60+10} = \frac{50}{70} = 0.714$$

Example 3. Annual income of two groups of families are given below:

Group I (₹)	Group II (₹)
26800	25200
26800	26000
26800	26800
26800	27600
26800	28400

Calculate range and coefficient of range.

Solution Range = L-S

Here For Group I: L = 26800 and S = 26800

Group I: L - S = 26800-26800 = 0

Here For Group II: L = 28400 and S = 25200

Group II: L - S = 28400-25200 = 3200

$$\text{Coefficient of range} = \frac{L-S}{L+S}$$

Group I: $\text{Coefficient of range} = \frac{26800-26800}{26800+26800} = \frac{0}{531600} = 0$

Group II: $\text{Coefficient of range} = \frac{28400-25200}{28400+25200} = \frac{3200}{53600} = 0.0597$

Example 4. Calculate the coefficient of range for the following data:

Wages	50-60	60-70	70-80	80-90	90-100	100-110	110-120
No. of Labourers	60	69	70	72	80	82	88

Solution $\text{Coefficient of range} = \frac{L-S}{L+S}$

Here: L = 26800 and S = 26800

$$\text{Coefficient of range} = \frac{120-60}{120+60} = \frac{60}{180} = 0.333$$

1.35 Inter-quartile Range and Quartile Deviation

Inter-quartile range and quartile deviations are the measures of range of parts. The range is based on two extreme values while quartile deviation is based on the middle fifty percent of the distribution.

Inter-quartile range is defined as the difference between the upper and lower quartiles.

$$\text{Inter-quartile range} = Q_3 - Q_1$$

Quartile deviation (Q.D.) is defined as the half of the inter-quartile range and also ca semi inter-quartile range.

$$\text{Quartile Deviation (Q.D.)} = \frac{Q_3 - Q_1}{2}$$

It is an absolute measure of dispersion.

The coefficient of quartile deviation is given by $\frac{Q_3 - Q_1}{Q_3 + Q_1}$.

1.35.1 Advantages and disadvantages of Quartile Deviation

Advantages

1. It is not affected by extreme values.
2. It is simple to understand and easy to calculate
3. It can be computed if the distribution is open ended.
4. It can be calculated even if the class intervals are unequal.
5. It is also useful where extreme values are likely to affect the results.

Disadvantages

1. It is affected by sampling fluctuation
2. It is not suitable for further mathematical treatment.
3. It is not based on all the values as it ignores half the items – first 25% and the last 25%

Illustrative Examples

Example 1. Find out quartile deviation and coefficient of quartile deviation from the followi scholarships of the students of various courses; 155, 140, 210, 220, 208, 244, 16 170, 188, 267, 270, 310, 355, 370 and 400

Solution Arranging the given data in ascending order as:

140, 155, 169, 170, 188, 208, 210, 220, 244, 267, 270, 310, 355, 370, 400.

Here, $$Q_1 = \frac{(N+1)^{th}\,item}{4} = \frac{(15+1)^{th}\,item}{4} = 4^{th}\,item = 170$$

and $$Q_3 = \frac{3(N+1)^{th}\,item}{4} = \frac{3(11+1)^{th}\,item}{4} = 12^{th}\,item = 310$$

Inter-quartile range $= Q_3 - Q_1 = 310 - 170 = 140$

Quartile deviation (Q.D.) $= \frac{Q_3 - Q_1}{2}$

$$= \frac{140}{2} = 70$$

and Coefficient of Q.D. $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$$= \frac{310 - 170}{310 + 170} = \frac{140}{480} = 0.292$$

Example 2. Consider the following table containing the heights of students in inches. Compute quartile deviation, coefficient of quartile deviation.

Marks	58	59	60	61	62	63	64	65	66	67	68
No. of Students	1	3	6	15	10	5	4	3	1	1	0

Solution

Marks	**Frequency (No. of Students)**	**Cumulative frequency**
58	1	1
59	3	4
60	6	10
61	15	25
62	10	35
63	5	40
64	4	44
65	3	47
66	1	48
67	1	49
68	0	49

$$Q_1 = \frac{(N+1)^{th}\,item}{4} = \frac{(49+1)^{th}\,item}{4}$$

$$= 12.5^{th}\,item = 61 \text{ inches}$$

and $$Q_3 = \frac{3(N+1)^{th}\,item}{4} = \frac{3(49+1)^{th}\,item}{4}$$

$$= 37.57^{th}\,item = 63 \text{ inches}$$

Inter-quartile range $= Q_3 - Q_1$

$$= 63 - 61 = 2 \text{ inches}$$

Quartile deviation (Q.D.) $= \dfrac{Q_3 - Q_1}{2}$

$$= \frac{2}{2} = 1.0$$

and Coefficient of Q.D. $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

$$= \frac{2}{63+61} = \frac{2}{124} = 0.0161$$

Example 3. Calculate quartile deviation and coefficient of quartile deviation from the following data:

Size	5-8	8-11	11-14	14-17	17-20
Frequency	14	24	38	20	4

Solution

Size	Frequency	Cumulative Frequency
5-8	14	14
8-11	24	38
11-14	38	76
14-17	20	96
17-20	4	100
	N = 100	

Here, $\frac{N}{4} = \frac{100}{4} = 25$ and $\frac{3N}{4} = \frac{3 \times 100}{4} = 75$

Q_1 lies in the interval 8-11

$\therefore \quad l = 8, \frac{N}{4} = 25, c.f. = 14, f = 24, h = 3$

$$Q_1 = l + \frac{h}{f}\left(\frac{N}{4} - c.f.\right)$$

$$= 8 + \frac{3}{24}(25 - 14) = 8 + 1.375 = 9.375$$

Q_3 lies in the interval 11-14

$l = 11, \frac{3N}{4} = 75, c.f. = 38, f = 38, h = 3$

$$Q_3 = l + \frac{h}{f}\left(\frac{3N}{4} - c.f.\right)$$

$$= 11 + \frac{75 - 38}{38} \times 3 = 11 + 2.921 = 13.921$$

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

$$= \frac{13.921 - 9.375}{2} = 2.273$$

and Coefficient of $\text{Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{13.921 - 9.375}{13.921 + 9.375} = \frac{4.546}{23.296} = 0.195$

Example 4. Consider the following table containing monthly income of Mr. X during the year. Compute quartile deviation, coefficient of quartile deviation.

Months	Monthly Income (x) (₹ in thousands)
1	39
2	40
3	40
4	41
5	41
6	42
7	42
8	43
9	43
10	44
11	44
12	45

Solution Here, $Q_1 = \frac{(N+1)^{th}\,item}{4} = \frac{(12+1)^{th}\,item}{4} = 3.25^{th}\,item$

$= 3^{rd}\,item + 25\%\ of\,(4^{th}\,item - 3^{rd}\,item)$

$= 40 + 0.25\,(4 - 40) = 40 + 0.25 = 40.25$

and $Q_3 = \frac{3(N+1)^{th}\,item}{4} = \frac{3(12+1)^{th}\,item}{4} = 9.75^{th}\,item$

$= 9^{th}\,item + 75\%\ of\,(10^{th}\,item - 9^{th}\,item)$

$= 43 + 0.75(44 - 43)$

$= 43 + 0.75 = 43.75$

Inter-quartile range $= Q_3 - Q_1$

$= 43.75 - 40.25 = 3.50$

Quartile deviation (Q.D.) $= \frac{Q_3 - Q_1}{2}$

$= \frac{3.50}{2} = 1.75$

and Coefficient of Q.D. $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{3.50}{43.75 + 40.25} = \frac{3.50}{84} = 0.0416$

and Coefficient of Q.D. $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

$$= \frac{56.17 - 38.94}{56.17 + 38.94}$$

$$= \frac{17.23}{95.11} = 0.181$$

1.36 Mean Deviation or Average Deviation

Mean deviation (M.D.) is defined as the average of the absolute deviations taken from an average usually, the mean, median or mode. Mean deviation is a measure of dispersion which is based on all values of asset of data. Mean deviation is based on all the items of the distribution and is calculated as an average, on the basis of deviation obtained from mean, median or mode – generally from the median.

1.36.1 Mean Deviation from Ungrouped Data

The formula for calculating mean deviation (M.D.) of n values $x_1, x_2, \ldots, x_n$ is

$$\text{M.D.} = \frac{1}{n}\sum_{i=1}^{n} |x_i - A| \ ,$$

where, A = Any constant out of mean, median and mode.

1.36.2 Mean Deviation from Grouped Data by Direct Method

For a frequency distribution in which the variate value x_i occurs f_i times $(i = 1, 2, \ldots, k)$, the formula is

$$\text{M.D.} = \frac{1}{N}\sum_{i=1}^{k} f_i |x_i - A|$$

where, $N = \sum_{i=1}^{k} f_i$ and A as defined above.

1.36.3 Mean Deviation from Grouped Data by Short-cut Method

When mean (or median) is not a whole number but a fraction value then the computation of mean deviation is quite complicated. In this case we use the following formula:

$$\text{M.D.} = \frac{\sum f|dx_i| + \left(\sum f_B - \sum f_A\right)(A - \bar{x})}{N}$$

where, $|dx| =$ Deviation from assumed mean

$\sum f_B$ = Total of frequencies above the mean

$\sum f_A$ = Total of frequencies below the mean

A = Assumed mean

$\bar{x}$ = actual mean

1.36.4 Coefficient of Mean Deviation

Coefficient of Mean Deviation is a relative measure of dispersion and is calculated by

$$\text{Coefficient of M.D.} = \frac{Mean\ Deviation}{Average\ used\ (A)}$$

1.36.5 Advantages and Disadvantages of Mean Deviation

Advantages

1. It is based on all values.
2. It is least affected by extreme values.
3. It is simple to understand and easy to calculate.

Disadvantages

1. It is not suitable for further mathematical treatment.
2. It cannot be calculated if the distribution is open ended.
3. The foremost weakness of M.D. is that in its calculation negative differences are tak positive without any sound reasoning.

Illustrative Examples

Example 1. Calculate the mean deviation from mean and median for the following data relating the students securing marks in a subject: 30, 33, 34, 35, 37, 39, 41, 45, 48.

Solution Here, $n = 9$

$$\text{A.M.}(\bar{x}) = \frac{1}{9}\sum_{i=1}^{9} x_i$$

$$= \frac{30+33+34+35+37+39+41+45+48}{9}$$

$$= \frac{342}{9} = 38$$

Here, $Q_1 = \frac{(N+1)^{th}\ item}{2} = \frac{(9+1)^{th}\ item}{2} = 5^{th}\ item = 37$

$\therefore$ Median = 37

Marks (x_i)	$\lvert x_i - mean \rvert$	$\lvert x_i - median \rvert$
30	8	7
33	5	4
34	4	3
35	3	2
37	1	0
39	1	2
41	3	4
45	7	8
48	10	11
	$\sum \lvert x_i - mean \rvert = 42$	$\sum \lvert x_i - median \rvert = 41$

$$\text{Mean deviation from mean} = \frac{\sum \lvert x - Mean \rvert}{N} = \frac{42}{9} = 4.67$$

$$\text{Mean deviation from median} = \frac{\sum \lvert x - Median \rvert}{N} = \frac{41}{9} = 4.56$$

Example 2. Calculate the mean deviation from mean and median for the following data:

10, 20, 42, 48, 50, 55, 60, 70, 80, 95

Solution

x_i	$\lvert x_i - median \rvert$	x_i	$\lvert x_i - mean \rvert$
10	42.5	10	43
20	32.5	20	33
42	10.5	42	11
48	4.5	48	5
50	2.5	50	3
55	2.5	55	2
60	7.5	60	7
70	17.5	70	17
80	27.5	80	27
95	42.5	95	42
$\sum \lvert x_i - median \rvert = 190$		$\sum x = 530$	$\sum \lvert x_i - mean \rvert = 190$

Here, $\left(\frac{n+1}{2}\right)^{th} term = \left(\frac{10+1}{2}\right)^{th} term = 5.5^{th} term$

$\therefore$ Median = 52.5

$$\text{Mean deviation from median} = \frac{\sum |x - Median|}{N}$$

$$= \frac{190}{10} = 19$$

$$\text{Coefficient of median deviation} = \frac{MD}{Median}$$

$$= \frac{19}{52.5} = 0.36$$

$$\text{Here, Mean } \bar{x} = \frac{\sum x}{N} = \frac{530}{10} = 53$$

$\therefore$ Mean = 53

$$\text{Mean deviation from mean} = \frac{\sum |x - Mean|}{N}$$

$$= \frac{190}{10} = 19$$

$$\text{Coefficient of median deviation} = \frac{MD}{Mean} = \frac{19}{53} = 0.35$$

$$= \frac{19}{53} = 0.35$$

Example 3. Calculate the mean deviation from mean for the following data:

Class	3-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	18	16	15	12	10	5	2	2

Solution

Class	Mid Value x_i	Frequency f_i	$dx_i = x_i - A$	$f_i dx_i$
0-10	5	18	-30	-540
10-20	15	16	-20	-320
20-30	25	15	-10	-150
30-40	35 (A)	12	0	0
40-50	45	10	10	100
50-60	55	5	20	100
60-70	65	2	30	60
70-80	75	2	40	80
Total		80		-670 and 1350 (ignoring signs)

Let A = Assumed mean = 35

$$\text{Mean}(\bar{x}) = A + \frac{\sum fdx}{N}$$

$$= 35 + \frac{(-670)}{80}$$

$$= 35 - 8.375 = 26.625$$

Now, Mean deviation from mean $= \dfrac{\sum f|dx| + \left(\sum f_B - \sum f_A\right)(A - \bar{x})}{N}$

where, $\sum f_B = 18 + 16 + 15 = 49$ (Total of frequencies above the mean)

$\sum f_A = 12 + 10 + 5 + 2 + 2 = 31$ (Total of frequencies below the mean)

$\sum f|dx| = 1350$ (Ignoring signs of dx)

$\therefore$ Mean deviation from mean $= \dfrac{1350 + (49 - 31) \times (26.625 - 35)}{80}$

$$= \frac{1350 + 18 \times (-8.375)}{80}$$

$$= \frac{1199.25}{80} = 14.99 \approx 15$$

Coefficient of mean deviation $= \dfrac{\textit{Mean Deviation}}{\bar{x}}$

$$= \frac{15}{26.625} = 0.563$$

Example 4. Calculate the mean deviation from mean for the following data:

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	6	5	8	14	6	2

Solution

Marks	**Frequency** f_i	**Mid Value** x_i	$f_i x_i$	$\lvert x_i - \bar{x}\rvert$	$f_i \lvert x_i - \bar{x}\rvert$
0-10	6	5	30	23.6	141.6
10-20	5	15	75	13.6	68
20-30	8	25	200	3.6	28.8
30-40	14	35	490	6.4	89.6
40-50	6	45	270	16.4	98.4
50-60	2	55	110	26.4	52.8
	$N = 41$		$\sum f_i x_i = 1175$		$\sum f_i \lvert x_i - \bar{x}\rvert = 479.2$

$$\text{Mean } (\bar{x}) = \frac{\sum fx}{N} = \frac{1175}{41} = 28.6$$

$$\text{Mean deviation from mean} = \frac{\sum f|x - \bar{x}|}{N}$$

$$= \frac{479.2}{41} = 11.69$$

Example 5. Calculate the mean deviation from median and its coefficient for the following data

Marks	No. of Students	Marks	No. of Students
20-25	5	50-55	9
25-30	11	55-60	7
30-40	16	60-70	4
40-45	29	70-80	10
45-50	9		

Solution

Marks	f_i	Mid value x_i	c.f.	$\lvert x_i - Median \rvert$	$f_i \lvert x_i - Median \rvert$
20-25	5	22.5	5	20.6	103
25-30	11	27.5	16	15.6	171.6
30-40	16	35	32	8.1	129.6
40-45	29	42.5	61	0.6	17.4
45-50	9	47.5	70	4.4	39.6
50-55	9	52.5	79	9.4	84.6
55-60	7	57.5	86	14.4	100.8
60-70	4	65	90	21.9	87.6
70-80	10	75	100	31.9	319
	$N = 100$				1053.2

$$\text{Median} = \frac{N}{2} = \frac{100}{2} = 50$$

The median class is $40 - 45$

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

$$= 40 + \frac{5}{29}\left(\frac{100}{2} - 32\right)$$

$$= 43.103$$

Median = 43.10

$$\text{Mean deviation from median} = \frac{\sum f|x_i - Median|}{N}$$

$$= \frac{1053.2}{100} = 10.53$$

$$\text{Coefficient of Mean deviation} = \frac{Mean\ deviation}{Median}$$

$$= \frac{10.53}{43.10} = 0.244$$

Example 6. Calculate

i) range

ii) mean deviation from median and mean

from the following data: 9, 8, 5, 9, 1, 6, 8, 2.

Solution $\text{Mean} = \frac{9+8+5+9+1+6+8+2}{8} = \frac{48}{8} = 6$

For median, arranging the values in ascending order 1, 2, 5, 6, 8, 8, 9, 9

$$\text{Median} = \frac{4^{th}\ item + 5^{th}\ item}{2} = \frac{6+8}{2} = \frac{14}{2} = 7$$

(i) Range $= = L - S = 9 - 1 = 8$

(ii) Mean deviation $= \frac{1}{n}\sum_{i=1}^{n} |x_i - A|$

A = Any constant out of mean, median and mode

n = 8 (Number of observations)

x_i	$\|x_i - A\|$ A = mean (6)	$\|x_i - A\|$ A = median (7)
1	5	6
2	4	5
5	1	2
6	0	1
8	2	1
8	2	1
9	3	2
9	3	2
	$\sum\|x_i - 6\| = 20$	$\sum\|x_i - 7\| = 20$

$$\text{Now, M.D. from mean} = \frac{\sum|x_i - 6|}{8} = \frac{20}{8} = 2.5$$

$$\text{M.D. from median} = \frac{\sum|x_i - 7|}{8} = \frac{20}{8} = 2.5$$

1.37 Standard Deviation

The idea of standard deviation (S.D.) was given by Karl Pearson in 1893. Standard deviation is defined as the square root of the arithmetic mean of the square of the deviation measured from the mean. It is denoted by σ (sigma). Standard deviation is also known as mean error, mean square error or root mean square deviation from mean.

1.37.1 Standard Deviation from Ungrouped Data by Direct Method

The formula for calculating standard deviation (S.D.) of n values $x_1, x_2, \ldots, x_n$ is

$$\text{S.D. } (\sigma) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

When mean is not a whole number but a fractional value then the computation is quite complicated. In this case, we use the following formula:

$$\text{S.D. } (\sigma) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2}$$

or

$$\text{S.D. } (\sigma) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \left(\frac{\sum_{i=1}^{n}x_i}{n}\right)^2}$$

1.37.2 Standard Deviation from Ungrouped Data by Short-cut Method

The formula for calculating S.D. is

$$\text{S.D. } (\sigma) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{dx_i^2}{n} - \left(\frac{dx_i}{n}\right)^2}$$

or

$$\text{S.D. } (\sigma) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{dx_i^2}{n} - (\bar{x} - A)^2}$$

where $dx_i = x_i - A \ (i = 1, 2, \ldots, k)$;

A = assumed mean

1.37.3 Standard Deviation from Grouped Data by Direct Method

For a frequency distribution in which the variate value x_i occurs f_i time $(i = 1, 2, \ldots, k)$, the formula is

$$\text{S.D. } (\sigma) = \sqrt{\frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{N}}$$

or $$\text{S.D. } (\sigma) = \sqrt{\frac{\sum_{i=1}^{k} f_i x_i^2}{N} - \bar{x}^2}$$

$$= \sqrt{\frac{\sum_{i=1}^{k} f_i x_i^2}{N} - \left(\frac{\sum_{i=1}^{k} f_i x_i}{N}\right)^2}$$

where $N = \sum_{i=1}^{k} f_i$

1.37.4 Standard Deviation from Grouped Data by Short-cut Method

The formula for calculating S.D. is

$$\text{S.D. } (\sigma) = \sqrt{\frac{\sum_{i=1}^{k} f_i dx_i^2}{N} - \left(\frac{\sum_{i=1}^{k} f_i dx_i}{N}\right)^2}$$

or $$\text{S.D. } (\sigma) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \frac{dx_i^2}{n} - (\bar{x} - A)^2}$$

where $d_i = x_i - A \ (i = 1, 2, \ldots, k)$;

A = assumed mean

1.37.5 Standard Deviation by Step Deviation Method

In order to simplify further the calculation work, some common factor may be taken out from all the deviations. The formula is

$$\text{S.D. } (\sigma) = C \times \sqrt{\frac{\sum_{i=1}^{k} f_i dx_i'^2}{N} - \left(\frac{\sum_{i=1}^{k} f_i dx_i'}{N}\right)^2}$$

where $dx_i = \frac{x_i - A}{C}$ $(i = 1, 2, \ldots, k)$;

A = assumed mean and

C = common factor

1.37.6 Combined Standard Deviation

If two sets contain n_1 and n_2 observations with means $\bar{x}_1$ and $\bar{x}_2$ and standard deviations σ_1 and σ_2 respectively, then the standard deviation of the combined set is given by

$$\text{Combined standard deviation } (\sigma_{12}) = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$$

Where $d_1 = \bar{x}_{12} - \bar{x}_1$, $d_2 = \bar{x}_{12} - \bar{x}_2$

and $\bar{x}_{12} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

Note: 1. Standard deviation is independent of change of origin and dependent of change of scale.

2. Standard deviation is always positive.

1.37.7 Advantages and disadvantages of Standard Deviation

Advantages

1. It is rigidly defined.
2. It is based on all the values.
3. It is used for further mathematical treatments.
4. It is based on arithmetic mean, therefore, it has all the advantages of that measure.

Disadvantages

1. It is affected by extreme values
2. It is not easy to understand and not easy to calculate.
3. It cannot be used for comparing the dispersion of two or more series of observations given in different units.

1.38 Variance

The square of standard deviation i.e., σ^2 is known as variance.

$$\text{Variance} = (S.D.)^2$$

For finding the value of variance just takes the square of S.D. in all above formulas.

1.38.1 Standard Error of the Mean

The standard error of the mean is the standard deviation of the sample mean estimate of a population mean. Standard error of mean is usually estimated by the sample estimate of the population standard deviation divided by the square root of the sample size (assuming statistical independence of the values in the sample).

The standard error of the mean is given by

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

where

σ = Standard deviation of the population

n = Sample size (number of items in the sample)

In case of a frequency distribution, the standard error of the mean is given by

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$$

where

σ = Standard deviation of the population

$N = \sum f$

1.39 Coefficient of Variation

This relative measurement is developed by Karl Pearson. It is most popularly used to measure relative variation of two or more than two series. It shows the relationship between the standard deviation and the arithmetic mean expressed in terms of percentage. This measure is used to compare uniformity, consistency and variability in two different series. Coefficient of variation (C.V.) is a relative measure of dispersion and has no units. Coefficient of variation is given by:

$$\text{Coefficient of variation (C.V.)} = \frac{\textit{Standard Deviation}}{\textit{Mean}} \times 100$$

$$= \frac{\sigma}{\bar{x}} \times 100$$

The series having less C.V. is said to be less variable or more consistent, more uniform, more stable or more homogeneous. The series having more C.V. is said to be more variable or less consistent, less uniform, less stable or less homogeneous.

The main advantage of the C.V. is that the consistency, uniformity, equitability, ... homogeneity of entirely different series, measured in different units can be compared.

Illustrative Examples

Example 1. Find the standard deviation of 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2.

Solution Here, mean $\bar{x} = 2$, $n = 15$, and $\sum_{i=1}^{15}(x_i - \bar{x})^2 = 0$

$$\text{Standard deviation } (\sigma) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$= \sqrt{\frac{1}{15} \times 0} = 0$$

Example 2. Find the standard deviation if the sums of squares of deviations taken from 40 of 10 values is 810.

Solution Here, mean $\bar{x} = 40$, $n = 10$, and $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 810$

$$\text{Standard deviation } (\sigma) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$= \sqrt{\frac{1}{10} \times 810} = \sqrt{81} = 9$$

Standard Deviation $(\sigma) = 6$

Example 3. Find the coefficient of variation if the sum of squares of deviations taken from mean 40 of 10 values is 360.

Solution Here, mean $\bar{x} = 40$, $n = 10$, and $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 360$

$$\text{Standard deviation } (\sigma) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$= \sqrt{\frac{1}{10} \times 360} = \sqrt{36} = 6$$

$$\text{Coefficient of variation (C.V.)} = \frac{S.D.}{Mean} \times 100$$

$$= \frac{6}{40} \times 100 = 15$$

Example 4. Find the variance and the standard deviation and standard error of mean of the following data: 1, 2, 3, 4, 5, 67, 8, 9, 10

Solution Mean $(\bar{x}) = \frac{\sum x}{n} = \frac{1}{5} \times (1+2+3+4+5+6+7+8+9+10)$

$$= \frac{55}{10} = 5.5$$

Variance $(\sigma^2) = \frac{1}{n}\sum (x_i - \bar{x})^2$

$$= \frac{1}{10}\{(1-5.5)^2 + (2-5.5)^2 + (2-5.5)^2 + (4-5.5)^2 + (5-5.5)^2$$

$$+ (6-5.5)^2 + (7-5.5)^2 + (8-5.5)^2 + (9-5.5)^2 + (10-5.5)^2\}$$

$$= \frac{1}{10}\left[(-4.5)^2 + (-3.5)^2 + (-2.5)^2 + (-1.5)^2 + (-0.5)^2 + 1.5^2 + 2.5^2 + 3.5^2 + 4.5^2\right]$$

$$= \frac{1}{10}[20.25 + 12.25 + 6.25 + 2.25 + 0.25 + 2.25 + 6.25 + 12.25 + 20.25]$$

$$= \frac{1}{10}(82.5) = 8.25$$

Variance $(\sigma^2) = 8.25$

Standard Deviation $(\sigma) = \sqrt{8.25} = 2.87$

Example 5. If the mean of 200 observations is 100 and their variance is 36, then find the sum of squares of all the observations.

Solution Since $\bar{x} = 100$

$$\Rightarrow \bar{x} = 100 = \frac{\sum x}{n}$$

$$\Rightarrow \sum x = n \times 100 = 200 \times 100 = 20000$$

Since variance = 36

$$\Rightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2$$

$$\Rightarrow 36 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (100)^2$$

$$\Rightarrow 36 + (100)^2 = \frac{1}{200}\sum_{i=1}^{n} x_i^2$$

$$\Rightarrow 10036 = \frac{1}{200}\sum_{i=1}^{n} x_i^2$$

$$\Rightarrow \sum_{i=1}^{n} x_i^2 = 10036 \times 200 = 2{,}007{,}200$$

Thus sum of squares is 2,007,200

Example 6. 250 girls and 150 boys gave a competition and the results so obtained are:

Girls	Boys
Mean = 73 Standard deviation = 6.4	Mean = 72 Standard deviation = 7.0

Solution Combined mean $(\bar{x}_{12}) = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

$$= \frac{250 \times 73 + 150 \times 72}{250 + 150} = 72.63$$

Now, $d_1 = \bar{x}_{12} - \bar{x}_1 = 73 - 72.63 = 0.37$ and $d_2 = \bar{x}_{12} - \bar{x}_2 = 72 - 72.63 = 0.$

Combined standard deviation $(\sigma_{12}) = \sqrt{\dfrac{n_1\sigma_1^2 + n_1 d_1^2 + n_2\sigma_2^2 + n_2 d_2^2}{n_1 + n_2}}$

$$= \sqrt{\frac{250(6.4)^2 + 150(7)^2 + 250(0.37)^2 + 150(0.}{400}}$$

$$= 6.65$$

Example 7. Find the variance and standard error of mean from the frequency distribution

x	2	4	6	8	10	12	14	16
f	4	4	5	15	8	5	4	6

Solution

x_i	f_i	$x_i f_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
2	4	8	-7	49	196
4	4	16	-5	25	100
6	5	30	-3	9	45
8	15	120	-1	1	15
10	8	80	1	1	8
12	5	60	3	9	45
14	4	56	5	25	100
16	6	80	7	49	245
	$N = 50$				$\sum f_i(x_i - \bar{x})^2 = 75$

$$N = \sum f_i = 50$$

$$\text{Mean } (\bar{x}) = \frac{\sum fx}{\sum f} = \frac{450}{50} = 9$$

$$\text{Variance } (\sigma^2) = \frac{\sum f_i (x_i - \bar{x})^2}{N} = \frac{754}{50} = 15.08$$

$$= \frac{754}{50} = 15.08$$

$$\text{Standard deviation } (\sigma) = \sqrt{Variance} = \sqrt{15.08} = 3.88$$

$$\text{Standard error of the mean } (\sigma_{\bar{x}}) = \frac{\sigma}{\sqrt{N}}$$

$$= \frac{3.88}{\sqrt{50}} = 0.5487$$

Example 8. Compute the standard deviation using step deviation method:

C.I.	20-30	30-40	40-50	50-60	60-70	70-80	80-90
f	3	6	13	15	14	5	4

Solution Let A = 55 (assumed mean) and C = 10

C.I.	f_i	x_i	$dx_i' = \frac{x_i - A}{C}$	$f_i dx_i'$	$f_i dx_i'^2$
20-30	3	25	-3	-9	27
30-40	6	35	-2	-12	24
40-50	13	45	-1	-13	13
50-60	15	55	0	0	0
60-70	14	65	1	14	14
70-80	5	75	2	10	20
80-90	4	85	3	12	36
	$N = 60$			$\sum f_i dx_i' = 2$	$\sum f_i dx_i'^2 = 134$

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

$$= 10 \times \sqrt{\frac{134}{60} - \left(\frac{2}{60}\right)^2} \qquad = 10 \times \sqrt{2.232} = 14.94$$

Example 9. Find the standard deviation and standard error of mean of the following data:

11, 14, 15, 17, 18.

Solution Mean $(\bar{x}) = \frac{\sum x}{n} = \frac{1}{5} \times (11+14+15+17+18) = \frac{75}{5} = 15$

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
11	-4	16
14	-1	1
15	0	0
17	2	4
18	3	9
		$\sum (x_i - \bar{x})^2 = 30$

Standard deviation $(\sigma) = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2}$

$$= \sqrt{\frac{1}{5} \times 30} = \sqrt{6} = 2.45$$

Standard error of the mean $(\sigma_{\bar{x}}) = \frac{\sigma}{\sqrt{n}}$

$$= \frac{2.45}{\sqrt{5}} = 1.096$$

Example 10. Find the standard deviation from the following frequency distribution by s method:

Size of item	6	7	8	9	10
Frequency	3	6	9	13	8

Solution Let $A = 8$ (assumed mean)

x_i	f_i	$dx_i = x_i - A$	$f_i dx_i$	$f_i dx_i^2$
6	3	-2	-6	12
7	6	-1	-6	6
8	9	0	0	0
9	13	1	13	13
10	8	2	16	32
11	5	3	15	45
	$N = \sum f_i = 44$		$\sum f_i dx_i = 32$	$\sum f_i dx_i^2$

S

Example 11.

Solution

$$\text{Standard deviation } (\sigma) = \sqrt{\frac{\sum f_i dx_i^2}{N} - \left(\frac{\sum f_i dx_i}{N}\right)^2}$$

$$= \sqrt{\frac{108}{44} - \left(\frac{32}{44}\right)^2}$$

$$= \sqrt{2.454 - 0.529} = \sqrt{1.925} = 1.387$$

ple 11. Find the standard deviation and standard error of mean of the following data:

80, 35, 45, 30, 70, 42, 36, 48, 90, 74

ion $\bar{x} = \frac{\sum x}{n} = \frac{1}{10} \times (80 + 25 + 45 + 30 + 70 + 42 + 36 + 48 + 90 + 74)$

$$= \frac{550}{10} = 55$$

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
80	25	625
35	-20	400
45	-10	100
30	-25	625
70	15	225
42	-13	169
36	-19	361
48	-7	49
90	35	1225
74	19	361
		$\sum (x_i - \bar{x})^2 = 4140$

$$\text{Standard deviation } (\sigma) = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2}$$

$$= \sqrt{\frac{1}{10} \times 4140} = \sqrt{414} = 20.35$$

$$\text{Standard error of the mean } (\sigma_{\bar{x}}) = \frac{\sigma}{\sqrt{n}}$$

$$= \frac{20.35}{\sqrt{10}} = 6.44$$

Example 12. Find the standard deviation and variance from the following frequency distribution.

Class Interval	0-8	8-16	16-24	24-32
Frequency	4	8	2	1

Solution Let $A = 12$ (assumed mean)

Class interval	f_i	x_i	$dx_i = x_i - A$	$f_i dx_i$	$f_i dx_i^2$
0-8	4	4	-8	-32	256
8-16	8	12 (A)	0	0	0
16-24	2	20	8	16	128
24-32	1	28	16	16	256
	$N = 15$			$\sum f_i dx_i = 0$	$\sum f_i dx_i^2 =$

$$\text{Standard deviation } (\sigma) = \sqrt{\frac{\sum f_i dx_i^2}{N} - \left(\frac{\sum f_i dx_i}{N}\right)^2}$$

$$= \sqrt{\frac{640}{15} - 0} = \sqrt{42.667} = 6.53$$

$$\text{Variance } (\sigma^2) = (6.53)^2 = 42.667$$

Example 13. The means of two samples of sizes 500 and 600 were 186 and 175 respectively. The standard deviations for the two samples were 9 and 10 respectively. Find the mean and variance of the combined sample.

Solution Here, $n_1 = 500$, $n_2 = 600$, $\bar{x}_1 = 186$, $\bar{x}_2 = 175$, $\sigma_1 = 9$ and $\sigma_2 = 10$

$$\text{Combined mean } (\bar{x}_{12}) = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

$$= \frac{500 \times 186 + 600 \times 175}{500 + 600}$$

$$= \frac{198000}{1100} = 180$$

Now, $d_1 = \bar{x}_{12} - \bar{x}_1 = 180 - 186 = -6$ and $d_2 = \bar{x}_{12} - \bar{x}_2 = 180 - 175 = 5$

$$\text{Combined standard deviation } (\sigma_{12}) = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$$

$$= \sqrt{\frac{500 \times [(9)^2 + (-6)^2] + 600 \times [(10)^2 + (5)^2]}{500 + 600}}$$

$$= \sqrt{\frac{585 + 750}{11}} = \sqrt{121.363} = 11.016$$

Example 14. Two samples of sizes 40 and 50 have the same mean 53 but different standard deviations 19 and 8 respectively. Find the standard deviation of the combined sample of size 90.

Solution Here, $n_1 = 40$, $n_2 = 50$, $\bar{x}_1 = \bar{x}_2 = 53$, $\sigma_1 = 19$ and $\sigma_2 = 8$

$$\text{Combined mean } (\bar{x}_{12}) = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

$$= \frac{40 \times 53 + 50 \times 53}{40 + 50}$$

$$= \frac{53 \times 90}{90} = 53$$

Now, $d_1 = \bar{x}_{12} - \bar{x}_1 = 53 - 53 = 0$ and $d_2 = \bar{x}_{12} - \bar{x}_2 = 53 - 53 = 0$

$$\text{Combined standard deviation } (\sigma_{12}) = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$$

$$= \sqrt{\frac{40 \times [(19)^2 + (0)^2] + 50 \times [(8)^2 + (0)^2]}{40 + 50}}$$

$$= \sqrt{\frac{17640}{90}} = \sqrt{196} = 14$$

Example 15. A person wants to invest ₹ 20,500 in one of the two companies A or B. Average return in a year from company A is ₹ 8,000 with a standard deviation of ₹ 25, while in company B, the average return in a year is ₹ 10,000 with a standard deviation of ₹ 40. Which company will you recommend to that person for investment?

Solution Coefficient of variation (C.V.) for company A $= \dfrac{S.D.}{Mean} \times 100$

$$= \frac{25}{8000} \times 100 = 0.3125$$

Coefficient of variation (C.V.) for company B $= \dfrac{S.D.}{Mean} \times 100$

$$= \frac{40}{10000} \times 100 = 0.40$$

Since C.V. for company A is less than the company B, company A is more consistent. Therefore, investment should be made in company A by that person.

Example 16. The following is the record of goals scored by team A in a football season:

No. of goals in match	0	1	2	3	4
No. of matches	1	9	7	5	3

For team B, the average number of goals scored per match was 2.5 with standard deviation of 1.25 goals. Find which team is more consistent.

Solution

For Team A

No. of goals (x_i)	No. of matches (f_i)	$f_i x_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
0	1	0	-2	4	256
1	9	9	-1	1	0
2	7	14	0	0	128
3	5	15	1	1	256
4	3	12	2	4	
	$N = \sum f_i = 25$	$\sum f_i x_i = 50$			$\sum f_i (x_i - \bar{x})^2 =$

$$\text{Mean } (\bar{x}) = \frac{\sum f_i x_i}{\sum f_i} = \frac{50}{25} = 2$$

$$\text{Standard deviation } (\sigma) = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}$$

$$= \sqrt{\frac{1}{25} \times 30} = \sqrt{1.2} = 1.095$$

$$\text{Coefficient of variation (C.V.)} = \frac{S.D.}{Mean} \times 100$$

$$= \frac{1.095}{2} \times 100 = 54.75$$

For Team B

Given, Mean $(\bar{x}) = 2.5$ and standard deviation $(\sigma) = 1.25$

$$\text{Coefficient of variation (C.V.)} = \frac{S.D.}{Mean} \times 100$$

$$= \frac{1.25}{2.5} \times 100 = 50$$

Since the coefficient of variation (C.V.) for team B is less than the team A. So, team B is more consistent than team A.

Example 17. The marks of a student in two subjects attempting the exams five times are as follows:

Subject A	Subject B
16	10
23	12
19	18
22	15
20	20

Find the subject in which the student has less variance and in which subject the student is more consistent.

Solution

Subject A

Subject A (x_i)	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
16	-4	16
23	3	9
19	-1	1
22	2	4
20	0	0
$\sum x_i = 100$		$\sum (x_i - \bar{x})^2 = 30$

$$\text{Mean } (\bar{x}) = \frac{\sum x_i}{n} = \frac{100}{5} = 20$$

$$\text{Standard deviation } (\sigma) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$= \sqrt{\frac{1}{5} \times 30} = \sqrt{6} = 2.45$$

$$\text{Variance } (\sigma^2) = (2.45)^2 = 6$$

Subject B

Subject B (y_i)	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
10	-5	25
12	-3	9
18	3	9
15	0	0
20	5	25
$\sum y_i = 75$		$\sum (y_i - \bar{y})^2 = 68$

$$\text{Mean } (\bar{y}) = \frac{\sum y_i}{n} = \frac{75}{5} = 15$$

$$\text{Standard deviation } (\sigma) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$= \sqrt{\frac{1}{5} \times 68} = \sqrt{13.6} = 3.69$$

$$\text{Variance } (\sigma^2) = (3.69)^2 = 13.60$$

$\Rightarrow$ The subject A has less variance than subject B.

$$\text{Coefficient of variation (C.V.) in subject A} = \frac{S.D.}{Mean} \times 100$$

$$= \frac{2.45}{20} \times 100 = 12.25$$

$$\text{Coefficient of variation (C.V.) subject B} = \frac{S.D.}{Mean} \times 100$$

$$= \frac{3.69}{15} \times 100 = 24.60$$

$\Rightarrow$ Since the coefficient of variation (C.V.) in subject A is less than subject
So, subject A has more stable marks.

Example 18. The mean and the standard deviation of 100 items are found to be 30
respectively. If at the time of calculation, two items were wrongly taken
47 instead of 23 and 26. Then find the correct mean and the correct
deviation. Also calculate the coefficient of variation