Unit-1.

Descriptive Statistics and methods for data science.

* Measures of central tendency

The importance of statastical analysis is to find a number which represents in some definite way the entire data. Such a representative number is called the central value (or) average. Hence, an average constitutes a measures of central, tendency of the series.

→ Measures of central tendency enable us to compare diffent groups of data.

The following are some important measures of central tendency in common use :-

(1) Arithmetic mean
(2) Median
(3) Mode
(4) Geometric mean
(5) Harmonic mean

① **Arithmetic mean** :- The arithmetic mean of a series of values of a variate is defined as the quantity obtained by dividing the sum of the values of variates by their number.

Ungrouped data :-

Let $x_1, x_2, x_3, \ldots\ldots, x_n$ be the n values of variate 'x' then

$$\boxed{AM = \frac{x_1+x_2+x_3+\cdots\cdots+x_n}{n} = \frac{\Sigma x_i}{n}}$$

4th Grouped Data.

Let $x_1, x_2, x_3, \dots, x_n$ be the mid values of n classes of a frequency distribution with frequencies $f_1, f_2, f_3, \dots, f_n$ respectively.

$$\boxed{AM = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{\Sigma x_i f_i}{N}}$$ $\because N = \Sigma f_i$ = total frequency.

② Median :-

Median of a distribution is defined as the value of variate which divides it into two equal parts. When the variates are arranged in ascending (or) descending order magnitude.

Ungrouped data

Arrange the variates in ascending / descending order magnitude. Determine the total number 'n' of variates.

If 'n' is odd then, median is the value of $\left(\frac{n+1}{2}\right)^{th}$ variate

If 'n' is even then, median is the average value of $\left(\frac{n}{2}^{th}, \frac{n}{2}+1^{th}\right)$ variates.

Grouped data

Calculate the less than cummulative frequency of the class. The class corresponding to the cummulative frequency just greater than $\frac{N}{2}$ is called the "median class."

$$\therefore \boxed{\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c}$$ [$\because$ c for class interval]

where,

l = lower limit of median class

N = total frequency

m = cummulative frequency of class preceeding the median class

f = frequency of median class

c/h = width / length of the median class.

③ **Mode :-**

Mode of distribution is defined as that value of vaeiate for which frequency is maximum.

Ungrouped data

When individual items are given mode is the value of item which occurs most frequently.

Grouped data

In a grouped data the class interval having the maximum frequency is called the modal class. Even, if the frequency distribution has class intervals of unequal magnitude the mode can be calculated provided the modal class and the class preceeding and succeding it are of the same magnitude.

$$\boxed{\text{Mode} = l + \frac{f - f_1}{2f - f_1 - f_2} \times c}$$

where,

l = lower limit of modal class

f = frequency of modal class

f_1 = frequency of class preceeding the modal class

f_2 = frequency of class succeeding the modal class.

c = width of modal class.

① Calculate the arithmetic mean from the following data.

Profit per shop	100-200	200-300	300-400	400-500	500-600	600-700	700-800
No. of Shops	10	18	20	26	30	28	18.

Sol:- Given that

Profits per shop	No. of shops (f)	Mid values (x)	fx
100-200	10	150	1500
200-300	18	250	4500
300-400	20	350	7000
400-500	26	450	11,700
500-600	30	550	16,500
600-700	28	650	18,200
700-800	18	750	13,500
	$\Sigma f = 150$		$\Sigma xf = 72,900$

$\therefore$ Mean $= \frac{\Sigma xf}{f} = \frac{72,900}{150} = 486.$

(or).

Using deviation method.

$d = \frac{x-A}{h} \Rightarrow \boxed{\text{mean} = A + \frac{h\Sigma fd}{N}}$

profit per shop	f	mid values (x)	$d = \frac{x-A}{h}$	fd
100-200	10	150	-3	-30
200-300	18	250	-2	-36
300-400	20	350	-1	-20
400-500	26	450 (A)	0	0
500-600	30	550	1	30
600-700	28	650	2	56
700-800	18	750	3	54
	$\Sigma f = 150$			$\Sigma fd = 54$

$$\text{Mean} = A + \frac{h\Sigma fd}{N}$$

$$= 450 + \frac{100(54)}{150}$$

$$= 450 + 36$$

$\therefore$ $\boxed{\text{Mean} = 486.}$

② Find the mean of the following 144, 148, 153, 156, 162.

Sol:- Given that

144, 148, 153, 156, 162.

$$\text{Mean} = \frac{\text{sum of values}}{\text{No. of values}}$$

$$= \frac{144+148+153+156+162}{5}$$

$$= \frac{773}{5}$$

$\therefore$ $\boxed{\text{Mean} = 154.60}$

③ Find the mean of the following data.

x	1	2	3	4	5	6
f	6	4	5	3	4	5

Sol:- Given that

x	f	xf
1	6	6
2	4	8
3	5	15
4	3	12
5	4	20
6	5	30
	$\Sigma f = 27$	$\Sigma xf = 91$

$$\text{Mean} = \frac{\Sigma xf}{\Sigma f} = \frac{91}{27}$$

$$\therefore \boxed{A_m = 3.372}$$

④ Find the mean of the following data.

Driving time	0-10	10-20	20-30	30-40	40-50
No. of teachers	3	10	6	4	2

Sol:- Given that

Driving time	(f) No. of teachers	Mid values (x)	xf
0-10	3	5	15
10-20	10	15	150
20-30	6	25	150
30-40	4	35	140
40-50	2	45	90
	$\Sigma f = 25$		$\Sigma xf = 545$

$$\text{Arithmetic mean} = \frac{\Sigma xf}{\Sigma f}$$

$$= \frac{545}{25}$$

$$\therefore \boxed{A_m = 21.8}.$$

⑤ Find the mean of the following data

weights	47	48	49	50	51
No. of students	6	4	2	2	1

Sol:- Given that,

weights (x)	No. of Students (f)	$d = \frac{x-A}{h}$	fd
47	6	-2	-12
48	4	-1	-4
49	2	0	0
50	2	1	2
51	1	2	2
	$\Sigma f = 15$		$\Sigma fd = -12$

$$\text{Mean} = A + \frac{h\Sigma fd}{N}$$

$$= 49 + \frac{1(-12)}{15}$$

$$= 49 - \frac{12}{15}$$

$$= 49 - \frac{4}{5}$$

$$= \frac{254-4}{5} = 48.2$$

$$\boxed{\text{Mean}(A_m) = 48.2}$$

⑥ Find the median of the following

57, 58, 61, 42, 38, 65, 72, 66.

Sol :- Given that

57, 58, 61, 42, 38, 65, 72, 66

Arrange in ascending order

38, 42, 57, 58, 61, 65, 66, 72.

Total no. of values = 8 (even)

Since n is even, the median is the average of $\frac{n}{2}$th & $(\frac{n}{2}+1)$th variate.

= $\frac{8}{2}$th and $(\frac{8}{2}+1)$th

= 4th & 5th term → average

i.e., $\frac{58+61}{2} = \frac{119}{2} = 59.5$

∴ $\boxed{\text{Median} = 59.5}$

* For discrete grouped data

Median is obtained by considering the cummulative frequencies.

Step-1 : Find '$\frac{N}{2}$'.

Step-2 : Identify the cummulative frequency just greater than $\frac{N}{2}$.

Step-3 : The value of x corresponding to the cummulative frequency just greater than $\frac{N}{2}$ is our required median.

① obtain the median for the following frequency distribution

x	1	2	3	4	5	6	7	8	9
f	8	10	11	16	20	25	15	9	6

Sol:- Given that

x	f	cf
1	8	8
2	10	18
3	11	29
4	16	45
5	20	65 $> \frac{N}{2}$
6	25	90
7	15	105
8	9	114
9	6	120
	$\Sigma f = 120$	

$N = \Sigma f = 120$

$\frac{N}{2} = \frac{120}{2} = 60$

cf is greater than is 65.

$\therefore$ $\boxed{\text{Median} = 5}$

* For continuous grouped data.

For continuous grouped data, the class corresponding to cummulative frequency just greater than $\frac{N}{2}$ is called Median class and the median is obtained by the following formula.

$$\boxed{\text{Median} = l + \frac{h}{f}\left[\frac{N}{2} - c\right]}$$

where, l = lower limit of median class

h = width/magnitude of median class

f = frequency of median class

N = total frequency

c = cummulative frequency preceeding the median class.

① Find the median of wages for the following:

Wages	2000-3000	3000-4000	4000-5000	5000-6000	6000-7000
No. of workers	3	5	20	10	5

Sol:- Given that,

wages	No. of workers(f)	cf
2000-3000	3	3
3000-4000	5	8 (c)
Median class $\boxed{4000-5000}$	20	28 $> \frac{N}{2}$
5000-6000	10	38
6000-7000	5	43
	$\Sigma f = 43$	

$\frac{N}{2} = \frac{43}{2} = 21.5$

$l = 4000$

$c = 8$

$h = 1000$

$f = 20.$

$N = 43$

$$\text{Median} = l + \frac{h}{f}\left[\frac{N}{2} - c\right]$$

$$= 4000 + \frac{1000}{20}\left[21.5 - 8\right]$$

$$= 4000 + 50\left[13.5\right]$$

$$= 4000 + 675$$

$\therefore \boxed{\text{Median} = 4675}$

(2) In a factory employing 3000 persons. In a day 5% work less than 3 hours, 580 work from 3.01 to 4.50 hrs, 30% work 4.51 to 5.00 hrs, 500 work from 6.01 to 7.50 hrs, 20% work from 7.51 to 9.00 hrs and at rest work 9.01 more hrs. What is the median hrs of the work.

Sol:- Given that

working hrs (CI)	f	cf	Modified CI
less than 3hrs	5% of 3000 = 150	150	less - 3.005
3.01 - 4.50	580	730 (c)	3.005 - 4.505
4.51 - 6.00	30% of 3000 = 900	1630 > $\frac{N}{2}$	$\boxed{4.505 - 6.005}$ median class
6.01 - 7.50	500	2130	6.005 - 7.505
7.51 - 9.00	20% of 3000 = 600	2730	7.505 - 9.005
more than 9.01	270	3000	9.005 - more
	$\Sigma f = 3000$		

$N = \Sigma f = 3000$

$\frac{N}{2} = \frac{3000}{2} = 1500.$

$l = 4.505$

$c = 730$

$f = 900$

$h = 1.5$

$$\text{Median} = l + \frac{h}{f}\left[\frac{N}{2} - c\right]$$

$$= 4.505 + \frac{1.5}{900}[1500 - 730]$$

$$= 4.505 + \frac{15}{900}(770)$$

$$= 4.505 + 1.2855$$

$$\therefore \boxed{\text{Median} = 5.7885}$$

① Find The mode of the following salaries.

(a) 850, 750, 600, 825, 850, 725, 600, 850, 640, 530.

(b) 40, 45, 48, 57, 78

(c) 45, 55, 50, 45, 40, 55, 45, 55.

Sol :- Given that

(a) 850, 750, 600, 825, 850, 725, 600, 850, 640, 530.

Most frequently occuring value is '850'.

∴ $\boxed{\text{Mode} = 850}$.

(b) 40, 45, 48, 57, 78

There is no repeated value.

∴ $\boxed{\text{Mode} = 0}$ (or) no mode.

(c) 45, 55, 50, 45, 40, 55, 45, 55

Here, 45 & 55 are most frequently repeated.

∴ $\boxed{\text{Mode} = 45, 55}$.

② Find the mode of the following distribution

CI	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
f	5	8	7	12	28	20	10	10

Sol :- Given that,

CI	f
0-10	5
10-20	8
20-30	7
30-40	12 f_1
M.C $\boxed{40-50}$	28 f
50-60	20 f_2
60-70	10
70-80	10

Here, $l = 40$

$f = 28 \Rightarrow 2f = 56.$

$f_1 = 12$

$f_2 = 20$

$c = 10$

$$\text{Mode} = l + \frac{f - f_1}{2f - f_1 - f_2} \times c$$

$$= 40 + \frac{28-12}{56-12-20} \times 10$$

$$= 40 + \frac{16}{24}(10)$$

$$= 40 + \frac{20}{3} = \frac{140}{3} = 46.6$$

$\therefore$ $\boxed{\text{Mode} = 46.6}$.

③ Find the mean, median, mode of the following distribution

CI	20-40	40-60	60-80	80-100	100-120	120-140	140-160	160-180	180-200
f	8	12	20	30	40	35	18	7	5

Sol:- Given that,

(i) mean.

CI	f	Mid values (x)	xf
20-40	8	30	240
40-60	12	50	600
60-80	20	70	1400
80-100	30	90	2700
100-120	40	110	4400
120-140	35	130	4550
140-160	18	150	2700
160-180	7	170	1190
180-200	5	190	950
	$\Sigma f_i = 175$		$\Sigma xf = 18,730$

$$\text{Mean} = \frac{\Sigma f_i x_i}{\Sigma f_i}$$

$$= \frac{181730}{175}$$

$$\boxed{\text{Mean} = 107.028}$$

(ii) Median.

CI	f	cf	
20-40	8	8	
40-60	12	20	
60-80	20	40	
80-100	30	70	(m)
MC $\boxed{100-120}$	40 (f)	110	$> \frac{N}{2}$
120-140	35	145	
140-160	18	163	
160-180	7	170	
180-200	5	175	
	$\Sigma f = 175$		

$N = \Sigma f = 175$

$\frac{N}{2} = \frac{175}{2} = 87.5$

$l = 100$, $m = 70$, $c = 20$

$\frac{N}{2} = 87.5$, $f = 40$

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$$

$$= 100 + \left(\frac{87.5 - 70}{\cancel{40}_{2}}\right)(\cancel{20})$$

$$= 100 + \frac{17.5}{2}$$

$$= 100 + 8.75$$

$$\therefore \boxed{\text{Median} = 108.75}$$

(iii) Mode :-

CI	f	
20-40	8	
40-60	12	
60-80	20	
80-100	30	f_1
Modal class $\boxed{100-120}$	40	f
120-140	35	f_2
140-160	18	
160-180	7	
180-200	5	
	$\Sigma f = 175$	

Here,

$l = 100$

$f = 40 \Rightarrow 2f = 80$

$f_1 = 30$

$f_2 = 35$.

$c = 20$.

$$\text{Mode} = l + \frac{f - f_1}{2f - f_1 - f_2} \times c$$

$$= 100 + \frac{40-30}{80-30-35} \times 20$$

$$= 100 + \frac{10}{15}(20)$$

$$= 100 + 13.33$$

$$\therefore \boxed{\text{Mode} = 113.33}$$

* **Geometric mean :-**

Geometric mean is defined as the n^{th} root of the product of the values of a distribution none of them being zero.

Geometric mean of an ungrouped data:

Let $x_1, x_2, x_3 \ldots\ldots x_n$ be the n values of the variate 'x' none of them being zero then, geometric mean 'G' is,

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \ldots\ldots x_n} = (x_1 \cdot x_2 \cdot x_3 \ldots\ldots x_n)^{1/n}$$

Take log on both sides

$\log G = \log (x_1 \cdot x_2 \cdot x_3 \cdots x_n)^{1/n}$

$= \frac{1}{n} \log (x_1 \cdot x_2 \cdot x_3 \cdots x_n)$

$= \frac{1}{n} [\log x_1 + \log x_2 + \cdots + \log x_n]$

$\log_{10} G = \frac{1}{n} \sum_{i=0}^{n} \log x_i$

$G = \text{Antilog} \left[\frac{1}{n} \sum_{i=0}^{n} \log x_i\right]$

$\therefore \boxed{G = 10^{\left(\frac{1}{n} \sum_{i=0}^{n} \log x_i\right)}}$

*** Geometric mean of grouped data :-**

Let $x_1, x_2, x_3, \ldots, x_n$ be the mid values of 'n' classes of a frequency distribution with frequencies $f_1, f_2, f_3, \ldots, f_n$

Let $N = \Sigma f_i$

The geometric mean is defined as the

Geometric mean $(G) = \sqrt[N]{x_1^{f_1} \cdot x_2^{f_2} \cdots x_n^{f_n}}$

$G = 10^{\frac{1}{N} \sum_{i=1}^{n} \log x_i^{f_i}}$

$\Rightarrow G = \text{Antilog} \frac{1}{N} \sum_{i=1}^{n} f_i \log x_i$

$\therefore \boxed{G = 10^{\left(\frac{1}{N} \sum_{i=1}^{n} f_i \log x_i\right)}}$

*** Harmonic mean :-**

The harmonic mean of a series of a values is defined as the reciprocal of the arithmetic mean of the reciprocal of individual items of which no item is equal to '0'

Harmonic mean of an ungrouped data :-

Let $x_1, x_2, x_3, \ldots, x_n$ be 'n' terms of the data there reciprocals are $\frac{1}{x_1}, \frac{1}{x_2}, \frac{1}{x_3}, \ldots, \frac{1}{x_n}$

Therefore, arithmetic mean of these reciprocals

$$= \frac{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \cdots + \frac{1}{x_n}}{n}$$

$$\boxed{A_m = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i}}$$

Harmonic mean = Reciprocals of the Am of the reciprocals

$$= \frac{1}{\frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i}}$$

$$\therefore \boxed{H_m = \frac{n}{\sum_{i=1}^{n} x_i}}$$

* Harmonic mean of grouped data :-

$$\text{Harmonic mean } (H_m) = \frac{N}{\sum_{i=1}^{n} \frac{f_i}{x_i}}$$

① Find the geometric mean of 2, 7, 5, 9, 11

sol:- Given values

2, 7, 5, 9, 11

Here, $n = 5$

$$G = \sqrt[5]{2(7)(9)(5)(11)} = \left((2)(7)(9)(5)(11)\right)^{1/5}$$

$$G = 10^{\left(\frac{1}{5} \sum_{i=0}^{5} \log x_i\right)}$$

x	$\log x$
2	0.3010
7	0.8451
5	0.6990
9	0.9542
11	1.0414
	$\sum \log x_i = 3.8407$

$$G = 10^{\frac{3.8407}{5}}$$

$$= 10^{0.7681}$$

$$\therefore \boxed{G = 5.8627}$$

② Find The geometric mean of the following data.

x	2	4	6	8
f	5	9	7	11

Sol:- Given that

x	f	$\log x$	$f \log x$
2	5	0.3010	1.505
4	9	0.6020	5.418
6	7	0.7781	5.4467
8	11	0.9031	9.9341
	$\Sigma f = 32$		22.3038

$$G = 10^{\frac{1}{N}\sum_{i=1}^{n} f_i \log x_i}$$

$$= 10^{\left(\frac{1}{32}\sum_{i=1}^{4} f_i \log x_i\right)}$$

$$= 10^{\frac{1}{32}(22.3038)}$$

$$= 10^{0.697}$$

$$\therefore \boxed{G = 4.977}$$

③ Find the geometric mean of the following distribution.

Humidity rating	60	62	64	68	70
No. of days	3	2	4	2	4

Sol:- Given that

x	f	$\log x$	$f \log x$
60	3	1.77815	5.33445
62	2	1.7924	3.5848
64	4	1.8062	7.2248
68	2	1.8325	3.665
	4	1.8451	7.3804

$\Sigma f = 15$, $\Sigma f \log x = 27.1894$.

Geometric mean $= \text{Antilog}\left(\frac{1}{N}\Sigma f_i \log x\right)$

$= \text{Antilog}\left(\frac{27.1894}{15}\right)$

$= \text{Antilog}(1.8126)$

$\boxed{GM = 64.9531}$.

(4) Find the geometric mean of the following distribution.

class	15-19	20-24	25-29	30-34	35-39	40-44
frequency	13	32	4	42	58	51

Sol:- ~~ele~~ Given that

class	f	mid value (x)	log x	f log x
15-19	13	17	1.2304	15.9952
20-24	32	22	1.3424	42.9568
25-29	4	27	1.4314	5.7256
30-34	42	32	1.5051	63.2142
35-39	58	37	1.5682	90.9556
40-44	51	42	1.6232	82.7832
	$\Sigma f = 200$			$\Sigma f \log x = 301.6306$

Geometric mean $(G_m) = \text{Antilog}\left[\frac{1}{N}\Sigma f_i \log x\right]$

$= \text{Antilog}\left[\frac{301.6306}{200}\right]$

$= \text{Antilog}[1.508153]$

$\therefore \boxed{G_m = 32.2218}$

⑤ Find the geometric mean of the following distribution:

class	0-2	2-4	4-6	6-8	8-10	10-12
frequency	5	16	13	7	5	4

Sol:- Given that,

class	f	mid value (x)	$\log x$	$f \log x$
0-2	5	1	0	0
2-4	16	3	0.47712	7.6336
4-6	13	5	0.6980	9.087
6-8	7	7	0.8451	5.9157
8-10	5	9	0.9542	4.771
10-12	4	11	1.0414	4.1656
	$\Sigma f = 50$			$\Sigma f \log x = 31.5729$

Geometric mean $= \text{Antilog}\left(\frac{1}{N} \Sigma f_i \log x\right)$

$= \text{Antilog}\left[\frac{31.5729}{50}\right]$

$= \text{Antilog}\,(0.631458)$

$\therefore \boxed{G_m = 4.28014}$

Harmonic mean :-

① Suppose a train moves 100 km with a speed of 40 km/hr then 150 km with a speed of 50 km/hr and next 135 km with a speed of 45 km/hr. Calculate the average speed.

Sol:- To get average speed we require harmonic mean of 40, 45, 50 with 100, 150, 135 as respectively frequency.

x	f	f/x
40	100	2.5
50	150	3
45	135	3
	$\Sigma f = 385$	$\Sigma f/x = 8.5$

$$\text{Harmonic mean} = \frac{\Sigma f}{\Sigma(f/x)}$$

$$= \frac{385}{8.5}$$

$$\boxed{H.M = 45.2941}$$

$\therefore$ The average speed is 45.29 km/hr.

② Find the harmonic mean of the following distribution.

class	0-2	2-4	4-6	6-8	8-10	10-12
frequency	5	16	13	7	5	4

Sol:- Given that

class	frequency(f)	midvalue(x)	f/x
0-2	5	1	5
2-4	16	3	5.3
4-6	13	5	2.6
6-8	7	7	1
8-10	5	9	0.55
10-12	4	11	0.36
	$\Sigma f = 50$		$\Sigma(f/x) = 14.81$

$$\text{Harmonic mean} = \frac{\Sigma f}{\Sigma(f/x)}$$

$$= \frac{50}{14.81}$$

$$\therefore \boxed{H.M = 3.3761}$$

③ Find the harmonic mean of the following distribution.

Marks	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Frequency	15	13	8	6	15	7	6

Sol:- Given that

Marks	frequency (f)	mid values (x)	f/x
30-40	15	35	0.4286
40-50	13	45	0.2888
50-60	8	55	0.1454
60-70	6	65	0.0923
70-80	15	75	0.2
80-90	7	85	0.0823
90-100	6	95	0.0631
	$\Sigma f = 70$		$\Sigma (f/x) = 1.3005$

$$\text{Harmonic mean} = \frac{\Sigma f}{\Sigma (f/x)}$$

$$= \frac{70}{1.3005}$$

$$\boxed{HM = 53.825}$$

④ Find the harmonic mean of the following distribution

Size of items	6	7	8	9	10	11
f	4	6	9	5	2	8

Sol:- Given that

x	f	f/x
6	4	0.66
7	6	0.8571
8	9	1.125
9	5	0.5556
10	2	0.2
	8	0.7273

$\Sigma f = 34$ $\quad\quad$ $\Sigma(f/x) = 4.1317.$

Harmonic mean $= \dfrac{\Sigma f}{\Sigma(f/x)}$

$= \dfrac{34}{4.1317}$

$\therefore \boxed{H.M = 8.2291}$

Measures of Dispersion :-

The measure which gives the idea of amount of scattering of the data around the central value is called the measure of dispersion.

There are four commonly used measures of dispersion, that are standard deviation, mean deviation, quartail deviation and range.

Standard deviation :-

If $x_1, x_2, x_3, \ldots, x_n$ be a set of 'n' observations forming a population then its standard deviation as

$$\boxed{S.D = \sigma = \left[\frac{1}{n}\Sigma(x_i - \bar{x})^2\right]^{1/2}} \rightarrow \text{raw data.}$$

where $\bar{x} = \dfrac{\Sigma x_i}{n}$

For simple frequency distribution

$$\boxed{S.D = \sigma = \left[\frac{1}{N}\Sigma f_i(x_i - \bar{x})^2\right]^{1/2}} \rightarrow \text{grouped data}$$

where, $N = \Sigma f_i$ (discrete).

Note :- 1, The square of standard deviation is known as variance.

2) If f is a real number standard deviation

$$\boxed{S.D = \left[\frac{1}{N}\Sigma f_i x^2 - (\bar{x})^2\right]^{1/2}} \quad \text{continuous}$$

(3) Coefficient of variance (CD) = $\frac{S.D (\sigma)}{\bar{x}} \times 100$.

(4) If the values of x and f (or) large then simplicity step deviation method can be employed in which the deviation of the given values of 'x' from any orbitary point 'A' is taken then,

$$\boxed{\sigma^2 = \frac{1}{N}\Sigma f d^2 - \left(\frac{1}{N}\Sigma f d\right)^2}$$ where, $d = x - A$.

Incase of grouped frequency, $d = \frac{x-A}{h}$

$$\boxed{\sigma^2 = \left[\frac{1}{N}\Sigma f d^2 - \left(\frac{1}{N}\Sigma f d\right)^2\right] \times h^2}$$

* Mean deviation (MD) :-

If $x_1, x_2, x_3, \ldots, x_n$ be the 'n' observations then mean deviation $\boxed{MD = \frac{1}{N}\Sigma|x-A|}$

where, A = mean (or) median (or) mode.

→ For grouped data, mean deviation

$$\boxed{MD = \frac{1}{N}\Sigma f|x-A|}$$

Range :-

It is the simplest measure the dispersion and it is defined as the difference b/w the greater and least value of the variable.

→ For grouped data

It is defined as the difference between the upper limit of the largest class and the lower limit of the smallest class.

① Find the range of the marks of the students given as

60, 72, 96, 28, 35, 10, 40, 9, 85, 25.

Sol:- Given marks

60, 72, 96, 28, 35, 10, 40, 9, 85, 25

Range = largest mark − lowest mark

$= 96 - 9$

$\boxed{\text{Range} = 87}$.

② Calculate the mean deviation of the variates

40, 62, 54, 68, 76 from arithmetic mean.

Sol:- Given data

40, 62, 54, 68, 76

$$\text{Mean} = \frac{40+62+54+68+76}{5} = 60$$

$$MD = \frac{1}{n} \Sigma |x_i - A|$$

$$= \frac{1}{5}[20+2+6+8+16]$$

$$= \frac{52}{5} = 10.4$$

$\therefore \boxed{MD = 10.4}$.

③ Find the mean deviation about the median for the following data.

6, 7, 10, 12, 13, 4, 12, 16.

Sol:- Given data

6, 7, 10, 12, 13, 4, 12, 16

Ascending order

$4, 6, 7, 10, 12, 12, 13, 16 \rightarrow \frac{10+12}{2} = \frac{22}{2} = 11.$

$$M.D = \frac{1}{n} \Sigma |x_i - A|$$

$$= \frac{1}{8} |5+4+1+1+2+7+1+5|$$

$$MD = \frac{26}{8} = 3.25$$

$\therefore \boxed{MD = 3.25}$.

(4) Find the mean deviation from the median for the following data.

x_i	6	9	3	12	15	13	21	22
f_i	4	5	3	2	5	4	4	3

Sol:- Given data

x	f	cf	$\lvert x-A\rvert$	$f\lvert x-A\rvert$
6	4	4	9	36
9	5	9	6	30
3	3	12	12	36
12	2	14	3	6
15	5	19 $> \frac{N}{2}$	0	0
13	4	23	2	8
21	4	27	7	28
22	3	30	8	24
	$\Sigma f = 30$			$\Sigma f\lvert x-A\rvert = 161$

$\frac{N}{2} = \frac{30}{2} = 15$

$\boxed{\text{Median}(A) = 15}$.

$\therefore MD = \frac{1}{N}\Sigma f\lvert x-A\rvert$

$= \frac{161}{30}$

$\boxed{MD = 5.3667}$

(5) Find the mean deviation from the mean for the following data

CI	0-100	100-200	200-300	300-400	400-500	500-600	600-700	700-800
f	4	8	9	10	7	5	4	3

Sol:- Given that

CI	f	x	$d = \frac{x-B}{h}$	fd	$\lvert x-A \rvert$	$f\lvert x-A \rvert$
0-100	4	50	-4	-16	308	1232
100-200	8	150	-3	-24	208	1664
200-300	9	250	-2	-18	108	972
300-400	10	350	-1	-10	8	80
400-500	7	450 (B)	0	0	92	644
500-600	5	550	1	5	192	960
600-700	4	650	2	8	292	1168
700-800	3	750	3	9	392	1176
	$\Sigma f = 50$			$\Sigma fd = -46$		$\Sigma f\lvert x-A \rvert = 7896$

$$\text{Mean}(A) = B + \frac{h \Sigma fd}{N}$$

$$= 450 + \frac{100(-46)}{50}$$

$$= 450 + (-92)$$

$$A = 358$$

$$M.D = \frac{1}{N} \Sigma f \lvert x - A \rvert$$

$$= \frac{1}{50} (7896)$$

$$\boxed{M.D = 157.92}$$

⑥ The following table gives the scales of 100 companies. Find the mean deviation from the mean.

Sales	40-50	50-60	60-70	70-80	80-90	90-100
No. of companies	5	15	25	30	20	5

CI	f	cf	Mid values (x)	|x-A|	f|x-A|
10-20	4	4	15	30	120
20-30	6	10	25	20	120
30-40	10	20 (c)	35	10	100
Median class $\boxed{40-50}$	20	$40 > \frac{N}{2}$	45	0	0
50-60	10	50	55	10	100
60-70	6	~~60~~ 56	65	20	120
70-80	4	60	75	30	120
	$\Sigma f = 60$				$\Sigma f\|x-A\| = 680$

$$\text{Median} = l + \frac{h}{f}\left[\frac{N}{2} - c\right]$$

$$\frac{N}{2} = \frac{60}{2} = 30.$$

$l = 40$ $\quad c = 20$

$f = 20$ $\quad h = 10$

$\frac{N}{2} = 30$

$$\text{Median}(A) = 40 + \frac{10}{20}[30-20]$$

$$= 40 + \frac{1}{2}(10)^{5}$$

$$\boxed{A = 45.}$$

$$\text{Mean deviation (M.D)} = \frac{1}{N}\Sigma f|x-A|$$

$$= \frac{680}{60}$$

$$\therefore \boxed{MD = 11.333}.$$

(9) Find the variance and standard deviation for the following data.

5, 12, 3, 18, 6, 8, 2, 10.

Sol:- Given data

5, 12, 3, 18, 6, 8, 2, 10.

$$\bar{x} = \frac{5+12+3+18+6+8+2+10}{8} = \frac{64}{8} = 8$$

$\boxed{\bar{x} = 8}$.

x_i	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$
5	-3	9
12	-4	16
3	-5	25
18	10	100
6	-2	4
8	0	0
2	-6	36
10	2	4
		$\Sigma(x_i-\bar{x})^2 = 194$

variance $(\sigma^2) = \frac{1}{n}\Sigma(x_i-\bar{x})^2 = \frac{194}{8} = 24.25$

Standard deviation (S.D) $= \sigma = \sqrt{24.25} = 4.924$.

(10) Find the standard deviation for the following data

CI	30-40	40-50	50-60	60-70	70-80	80-90	90-100
f	3	7	12	15	8	3	2

Sol:- Given that

CI	f	midvalues (x)	x^2	fx^2	fx
30-40	3	35	1225	3675	105
40-50	7	45	2025	14175	315
50-60	12	55	3025	36300	660
60-70	15	65	4225	63375	975
70-80	8	75	5625	45000	600
80-90	3	85	7225	21675	255
90-100	2	95	9025	18050	190
	$\Sigma f = 50$			$\Sigma fx^2 = 202250$	3100

$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{3100}{50} = 62$

$\boxed{\bar{x} = 62}$

Variance $(\sigma^2) = \frac{1}{N} \Sigma fx^2 - (\bar{x})^2$

$= \frac{1}{50} 202250 - 3844$

$= 4045 - 3844$

$\boxed{(\sigma^2) = 201}$

$S.D (\sigma) = \sqrt{201} = 14.177$

(11) Find the standard deviation for the following data by assumed mean.

CI	30-40	40-50	50-60	60-70	70-80	80-90	90-100
f	3	7	12	15	8	3	2

Sol:- Given that

CI	f	midvalue (x)	$d = \frac{x-A}{h}$	fd	fd^2	d^2
30-40	3	35	-3	-9	27	9
40-50	7	45	-2	-14	28	4
50-60	12	55	-1	-12	12	1
60-70	15	65 (A)	0	0	0	0
70-80	8	75	1	8	8	1
80-90	3	85	2	6	12	4
90-100	2	95	3	6	18	9
	$\Sigma f = 50$			$\Sigma fd = -15$	$\Sigma fd^2 = 105$	

$S.D = \frac{h}{N} \sqrt{N \Sigma fd^2 - (\Sigma fd)^2}$

$= \frac{10}{50} \sqrt{50(105) - 225}$

$= \frac{1}{5} \sqrt{5250 - 225} = \frac{1}{5} \sqrt{5025} = \frac{1}{5} (70.888)$

$\boxed{SD = 14.177}$

(12) Find the variance & standard deviation for the following data.

x	4	8	11	17	20	24	32
f	3	5	9	5	4	3	1

Sol:- Given that

x_i	f	xf	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	$f(x_i-\bar{x})^2$
4	3	12	~~-12~~ -10	100	300
8	5	40	-6	36	180
11	9	99	-3	9	81
17	5	85	3	9	45
20	4	80	6	36	144
24	3	72	10	100	300
32	1	32	18	324	324
	30	$\Sigma xf = 420$			$\Sigma f(x_i-\bar{x})^2 = 1374$

$$\bar{x} = \frac{420}{30} = 14.$$

$$\boxed{\bar{x} = 14}.$$

$$\text{Variance } (\sigma^2) = \frac{1}{N}\Sigma f(x_i-\bar{x})^2$$

$$= \frac{1}{30}(1374)$$

$$\boxed{\sigma^2 = 45.8}$$

$$\text{Standard deviation } (\sigma) = \sqrt{45.8}$$

$$\boxed{\sigma = 6.76}$$

~~(13) The scores o~~

Note:-

1) The distribution having greater coefficient of variation is said to be more variable than the other.

2) The distribution having lesser coefficient of variation is

Said to be more consistent than the other.

→ The series with lower value of standard deviation is said to be more conissient than the other series with greater standard deviation.

→ The series with greater standard deviation is called more dispersed than other.

① The scores of two crickters A and B in 10 innings are given here. Find who is a better run getter and who is more aconsistent player.

Scores of A	40	25	19	80	38	8	67	121	66	76
Scores of B	28	70	31	0	14	111	66	31	25	4

Sol:- Given that

x_i	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	y_i	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$
40	-14	196	28	-10	100
25	-29	841	70	~~48~~ 32	~~2304~~ 1024
19	-35	1225	31	-7	49
80	-26	676	0	-38	1444
38	-16	256	14	-24	576
8	-46	2116	111	73	5329
67	13	169	66	28	784
121	67	4489	31	-7	49
66	12	144	25	-13	169
76	22	484	4	-34	1156
$\Sigma x = 540$		10596	$\Sigma y = 380$		10680

$\bar{x} = \frac{\Sigma x}{n} = \frac{540}{10} = 54.$

$\bar{y} = \frac{\Sigma y}{n} = \frac{380}{10} = 38.$

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

$= \frac{1}{10}(10596)$

$(\sigma_x^2) = 1059.6.$

$\frac{\Sigma S.D}{\bar{x}} \times 100$

$\frac{\frac{1}{n}\Sigma(x_i - \bar{x})^2}{\bar{x}} \times 100$

$\frac{\frac{1}{10}(10596)}{540} \times 100$

$S.D = \sigma_x = \sqrt{1059.6} = 32.5515$

variance $(\sigma_y^2) = \frac{1}{n} \Sigma (y_i - y)^2$

$= \frac{1}{10}(10680)$

$(\sigma_y^2) = 1068.$

$S.D = \sigma_y = \sqrt{1068} = 32.6802.$

Coefficient of variance of A $= \frac{\sigma_x}{\bar{x}} \times 100$

$= \frac{32.5515}{54} \times 100$

$\boxed{C.V. \text{ of } A = 60.2814}$

Coefficient of variance of B $= \frac{\sigma_y}{\bar{y}} \times 100$

$= \frac{32.6802}{38} \times 100$

$\boxed{C.V. \text{ of } B = 86}$

C.V of A < C.V of B

∴ A is better run getter and a more consistent player.

② Lives of two models of refrigerators A and B are given below. Which refrigeration models would you suggest to purchase.

Life in years	0-2	2-4	4-6	6-8	8-10
Model A	5	16	13	7	5
Model B	2	7	12	19	9

Sol:- Given that

Life in years (CI)	Mid values (x_i)	x_i^2	Model A f_i	Model A $f_i x_i$	Model A $f_i x_i^2$	Model B f_i	Model B $f_i x_i$	Model B $f_i x_i^2$
0-2	1	1	5	5	5	2	2	2
2-4	3	9	16	48	144	7	21	63
4-6	5	25	13	65	325	12	60	300
6-8	7	49	7	49	343	19	133	931
8-10	9	81	5	45	405	9	81	729
			46	212	1222	49	297	2025

$$\text{Variance } (\sigma_A^2) = \frac{1}{N}\Sigma f_i x_i^2 - \left(\frac{\Sigma f_i x_i}{N}\right)^2$$

$$= \frac{1}{46}(1222) - \left(\frac{212}{46}\right)^2$$

$$= 26.565 - 21.2400$$

$$(\sigma_A^2) = 5.325.$$

Standard deviation $(\sigma_A) = \sqrt{5.325} = 2.3076.$

$$\text{variance } (\sigma_B^2) = \frac{1}{N}\Sigma f_i x_i^2 - \left(\frac{\Sigma f_i x_i}{N}\right)^2$$

$$= \frac{1}{49}(2025) - \left(\frac{297}{49}\right)^2$$

$$= 41.3265 - 36.7384$$

$$(\sigma_B^2) = 4.5881.$$

$$SD = \sigma_B = \sqrt{4.5881} = 2.142$$

$$C.V \text{ of } A = \frac{\sigma_A}{\bar{x}} \times 100$$

$$= \frac{2.3076}{4.609} \times 100$$

$$= 0.5007 \times 100$$

$$\boxed{C.V \text{ of } A = 50.067}$$

$$C.V \text{ of } B = \frac{\sigma_B}{\bar{x}} \times 100$$

$$= \frac{2.142}{6.061} \times 100$$

$$= 0.3534 \times 100$$

$$\boxed{C.V \text{ of } B = 35.340}$$

∴ Model A would ~~of~~ suggest to purchase. refrigerations

③ Goals scored by two teams A and B in footbal season are as follows :

No. of goals		0	1	2	3	4
No. of matches	A	24	9	8	5	4
	B	25	9	6	5	5

By calculating the standard deviation in each case. ~~In~~ Find which be consider more consistent

Sol :- Given that

x_i	A				B			
	f_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$f_i(x_i-\bar{x})^2$	f_i	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	$f_i(x_i-\bar{x})^2$
0	24	-1.12	1.2544	30.1056	25	-1.12	1.2544	31.36
1	9	-0.12	0.0144	0.1296	9	-0.12	0.0144	0.1296
2	8	0.88	0.7744	6.1952	6	0.88	0.7744	4.6464
3	5	1.88	3.5344	17.672	5	1.88	3.5344	17.672
4	4	2.88	8.2944	33.1776	5	2.88	8.2944	41.472
	$\Sigma f_i = 50$			87.28	$\Sigma f_i = 50$			95.28

$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f} = \frac{56}{50} = 1.12$, $\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f} = \frac{56}{50} = 1.12$

~~St~~ Variance of A $(\sigma_A^2) = \frac{1}{N}\Sigma f_i(x_i-\bar{x})^2$

$$= \frac{87.28}{50}$$

$$\sigma_A^2 = 1.7456$$

$$S.D = \sigma_A = \sqrt{1.7456} = 1.3212$$

Variance of B $(\sigma_B^2) = \frac{1}{N}\Sigma f_i(x_i-\bar{x})^2$

$$= \frac{95.28}{50}$$

$$(\sigma_B^2) = 1.9056$$

$$S.D = \sigma_B = \sqrt{1.9056} = 1.3804$$

Therefore team 'A' is more consistent.

* Quartile Deviation.

It is a location based measure of dispersion and is defined as

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

where, Q_1 = lower quartile

Q_3 = upper quartile

→ Quartile deviation is also known as "Semi-Inter Quartile Range", $Q_3 - Q_1$ is called 'Inter Quartile Range'.

where, $$\boxed{Q_1 = L + \frac{\frac{N}{4} - C}{f} \times h}$$

where, L = lower limit of class containing Q_1

f = frequency containing Q_1

h = width of the class

c = cummulative frequency of preceeding class

→ Cummulative frequency just greater than $\frac{N}{4}$ is the class containing Q_1.

→ Second Quartile is the median

→ Third Quartile, $$\boxed{Q_3 = L + \frac{\frac{3N}{4} - C}{f} \times h}$$

Q Calculate Quartile deviation of the following

Data	0-4	5-9	10-14	15-19	20-24	25-29	30-34
Frequency (f)	4	5	10	8	7	9	7

Sol:-

CI	f	class boundary	cf
0-4	4	-0.5-4.5	4
5-9	5	4.5-9.5	9 (c)
10-14	10 (f)	l $\boxed{9.5-14.5}$	19 $> \frac{N}{4}$
15-19	8	14.5-19.5	27
20-24	7	19.5-24.5	34
25-29	9	24.5-29.5	43 $> \frac{3N}{4}$
30-34	7	29.5-34.5	50
	$\Sigma f = 50$		

$\frac{N}{4} = \frac{50}{4} = 12.5$

Here, 9.5-14.5 → class containing 'Q_1';

$c = 9$

$l = 9.5$

$f = 10$

$h = 5$

$$Q_1 = L + \frac{\frac{N}{4} - c}{f} \times h$$

$$= 9.5 + \frac{12.5 - 9}{\cancel{10}\ 2} \times \cancel{5}$$

$$= 9.5 + \frac{3.5}{2}$$

$$= 9.5 + 1.75 = 11.25$$

$\frac{3N}{4} = 3(12.5)$

$= 37.5$

$\boxed{Q_1 = 11.25}$

$Q_3 =$

$$Q_3 = L + \frac{\frac{3N}{4} - c}{f} \times h \rightarrow$$

$L = \cancel{19.5}\ 24.5$

$f = 9$

$c = 34$

$h = 5$

$$= 24.5 + \frac{37.5 - 34}{9} \times 5$$

$$= 24.5 + \frac{3.5}{9} \times 5 \Rightarrow 24.5 + \frac{17.5}{9}$$

$= 24.5 + 1.944$

$\boxed{Q_3 = 26.44}$

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

$= \frac{26.44 - 11.25}{2}$

$= \frac{15.15}{2}$

$\therefore \boxed{Q.D = 7.575}$

② calculate the Quartile deviation of the wages from the following:

Weekly wages	35-36	36-37	37-38	38-39	39-40	40-41	41-42
No.of workers	14	20	42	54	45	18	7

Sol:- Given that

CI	f	cf
35-36	14	14
36-37	20	34 (c)
37-38	42 (f)	76 $> \frac{N}{4}$
38-39	54	130
39-40	45	175 (c)
40-41	18 (f)	293 $> \frac{3N}{4}$
41-42	7	300
	$\Sigma f = 300$	

$\frac{N}{4} = \frac{300}{4} = 75$

$l = 37$ $\quad h = 1$

$f = 42$

$c = 34$

$$Q_1 = L + \frac{\frac{N}{4} - C}{f} \times h$$

$$= 37 + \frac{75 - 34}{42} \times 1$$

$$= 37 + \frac{41}{42}$$

$$= 37 + 0.9761$$

$$\boxed{Q_1 = 37.9761}$$

$$\frac{3N}{4} = 3(75) = 225$$

$L = 40$

$C = 175$

$f = 18$

$h = 1$

$$Q_3 = L + \frac{\frac{3N}{4} - C}{f} \times h$$

$$= 40 + \frac{225 - 175}{18} \times 1$$

$$= 40 + \frac{50}{18}$$

$$= 40 + 2.777$$

$$\boxed{Q_3 = 42.777}$$

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

$$= \frac{\cancel{37.}\ 42.777 - 37.9761}{2}$$

$$= \frac{4.8017}{2}$$

$$\boxed{\therefore Q.D = 2.4}$$

* Skewness :-

The measures of central tendency and dispersion in adequate to characterise a distribution, completely may be suppoted by two variable measures : skewness & kurtosis.

→ A distribution which is not symmetrical is skewed distribution.

→ In such distribution mean, median, mode will not coincide.

→ The values are pulled apart.

Measures of the values are culled a part.

Test of Skewness :-

The absence of asymmetry (or) skewness can be stated under the following conditions.

(1) If the distribution is symmetry the following conditions are observed. The values of mean, mode & median are coincide.

(2) $(Q_3 - \text{median}) = (\text{median} - Q_1)$

(3) The sum of positive deviations is equal to the sum of negative deviations.

(4) The frequencies on either side of the mode are equal.

Similarly,

A skewed distribution have following chalacteristics.

(1) If the distribution is skewed the following conditions are observed.

(1) The values of mean, median, mode will not coincide.

(2) $(Q_3 - \text{median}) \neq (\text{median} - Q_1)$

(3) The sum of +ve deviations ≠ the sum of −ve deviations.

* Measures of skewness:-

Absolute skewness = mean − mode.

* If the value of mean > mode, the skewness is +ve.

* If the value of mean < mode, the skewness is -ve.

These are three impotant measures of relative skewness.

(1) Karlpeason's Coefficient of skewness

(2) Bowley's coefficient of skewness,

(3) kelly's coefficient of skewness.

* Generally, Karlpearson's method is widely used.

→ Karlpeason's coefficient of skewness

$$= \frac{\text{mean} - \text{mode}}{\text{Standard deviation}}$$

1) Calculate karlpearsons coefficient of skewness for the following data

v	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
f	2	5	7	13	21	16	8	3

Sol:- Given that

v	f	Midvalue (x)	$d = \frac{x-A}{h}$	fd	d^2	fd^2
0-5	2	2.5	-4	-8	16	32
5-10	5	7.5	-3	-15	9	45
10-15	7	12.5	-2	-14	4	28
15-20	13 (f_1)	17.5	-1	-13	1	13
$\boxed{20-25}$	21 (f)	22.5 (A)	0	0	0	0
25-30	16 (f_2)	27.5	1	16	1	16
30-35	8	32.5	2	16	4	32
35-40	3	37.5	3	9	9	27
	$\Sigma f = 75$			$\Sigma fd = -9$		$\Sigma fd^2 = 193$

$$\text{Mean} = A + \frac{h\Sigma fd}{N}$$

$$= 22.5 + \frac{5(-9)}{75}$$

$$= 22.5 - 0.6$$

$$\boxed{\text{Mean} = 21.9}$$

$$\text{Mode} = l + \frac{f - f_1}{2f - f_1 - f_2} \times c$$

$2f = 2(21) = 42$

$$= 20 + \frac{21 - 13}{42 - 13 - 16} \times 5$$

$$= 20 + \frac{8 \times 5}{13}$$

$$= 20 + \frac{40}{13}$$

$$= 20 + 3.077$$

$$\boxed{\text{Mode} = 23.077}$$

$$S.D = \frac{h}{N}\sqrt{N\Sigma fd^2 - (\Sigma fd)^2}$$

$$= \frac{5}{75}\sqrt{75(193) - (-9)^2}$$

$$= \frac{5}{75}\sqrt{14475 - 81}$$

$$= \frac{5}{75}\sqrt{14394}$$

$$= 0.066(119.975)$$

$$\boxed{S.D = 7.99 \approx 8}$$

$$\text{karl pearsons coefficient} = \frac{\text{mean} - \text{mode}}{SD}$$

$$= \frac{21.9 - 23.077}{8}$$

$$= \frac{-1.177}{8}$$

$$\therefore \boxed{K.C = -0.147}$$

(2) Calculate the Karl pearsons coefficient of skewness for the following 25, ~~50~~ 15, 23, 40, 27, 25, 23, 25, 20

Sol:- Given data

Mean $(\bar{x}) = \dfrac{25+15+23+40+27+25+23+25+20}{9}$

$\boxed{(\bar{x}) = 24.77.}$

$\boxed{\text{Mode} = 25}$.

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

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
25	0.23	0.0529
15	-9.77	95.4529
23	-1.77	3.1329
40	15.23	231.9529
27	2.23	4.973
25	0.23	0.0529
23	-1.77	3.1329
25	0.23	0.0529
20	-4.77	22.753
		361.5563

$(\sigma) = \sqrt{\frac{1}{9}(361.5563)}$

$\boxed{\sigma = 40.17}$

$K.C = \dfrac{\text{mean} - \text{mode}}{S.D}$

$= \dfrac{24.77-25}{40.17}$

$\therefore \boxed{K.C = -5.726}$

* Kurtosis :-

The expression kurtosis is used to describe the peakedness of the curve as far as the measurement of a shape is concerned. So, we have two characteristics.

→ Skewness which refers to assymetry of a series and Kurtosis which measures the peakedness of a normal curve.

→ This characteristic of frequency is termed as "kurtosis."

→ Measures of kurtosis denote the shape of the top of a frequency curve.

* Measures of kurtosis :-

→ Measures of kurtosis of a frequency distribution are based upon the fourth moment about the mean of the distribution.

→ Symmetrically $\beta_2 = \frac{\mu_4}{\mu_2^2}$ (or) $\frac{\mu_4}{s^4}$

where, μ_4 = fourth moment

μ_2 = second moment

→ If $\boxed{\beta_2 = 3}$, the distribution is said to be normal (the curve is 'mesokurtic'.

→ If $\boxed{\beta_2 > 3}$, the distribution is said to be more peaked and the curve is 'lepokurtic'

→ If $\boxed{\beta_2 < 3}$, the distribution is said to be flatter than normal curve and the curve is 'platikurtic'