BASIC CONCEPT OF STATISTICS

❖ Population

A group of all the units under the study is called a **population**
For example, to study the standard of living workers of factory, a group of all the workers of factory become a population for the study. The total number of units contain in the population is called the size of the population. If the total number of units of the population is finite say N call it finite population.

❖ Sample

A set or group of units selected from the population on the basis of some defined criteria is called a **sample**. The number of units in the sample is term as the sample size. For example, if we selected 150 workers from above population out workers by any statistical method then the group of selected workers is called a sample and the sample size n is equal to 150.

❖ Quantitative Data (Numerical variable)

If the variable characteristic is numerical then it is known as **the numeric variable** and the collection of the observations on the numerical variables it is called as the Quantatived data.
Ex.: - Income of workers, their age, weight, Hight.

❖ Qualitative Data (categorical variable)

If the variable characteristic is non numerical then it is called **Qualitative variable.** We shall call such qualitative characteristic as an attribute.
Ex.: - Gender of workers, their education level, feedback level.

❖ Discrete Data

if a variable can assume definite or countable values within the specific range, then it is called **discrete variable.**
Example: - Number of children per family, Number of accidents per day, number of printed errors per page in book etc example of discrete variables

❖ Continuous Data

If the variable which take all possible values within a given range and its value not countable than it is called **continuous data**

Example: - Height of the students, price of the product temperature of any place etc are the example of continuous data

❖ DESCRIPTIVE STATISTICS

Descriptive Statistics refers to those methods which are used for the collection, presentation as well as analysis of data. These methods relate measurement of central tendencies, measurement of dispersion, measurement of correlation etc. For Example: Descriptive statistics is used when you estimate average height of the secondary students in your school. Descriptive statistics is also used when you find the marks in science and mathematics of the students in all classes are intimately related to each other.

MEASURES OF CENTRAL TENDENCY

❖Measures of Central Tendency

The value of the variable is concentrated around a certain central value this characteristic of data is called as **central tendency**.

The central value around which values of the variables are concentrated is called as **measured of central tendency**. **OR** The value lies within the range of data it is known as a **measure of central tendency**.

❖Ideal or Good Measures of Central Tendency

This ideal measure is given by prof yules.

1) It should be well defined and rigid.
2) It should be based on all the observation of the data.
3) It should be stable measure. If means that values od averages found for different samples of same size from the same population should be almost same.
4) It should not be unduly affected by a few very large or very small observation.
5) It should be affected as little as possible by fluctuations of sampling.

❖ Name the Measures of Central Tendency

There are three the measures of central tendency.

i) Mean
ii) Median
iii) Mode

i) Mean (Arithmetic mean) (Average)

Arithmetic mean of a set of observations is their sum divided by the number of observations.

- **Merits: -**

a) It is rigidly defined.
b) It is easy to understand and easy to calculate.
c) It is based upon all the observations.
d) It is most suitable to algebraic treatment. The mean of the composite series in terms of the means and sizes of the component series.
e) Of all the averages, arithmetic mean is affected least by fluctuations of sampling. This property is sometimes described by saying that arithmetic mean is, a stable average

- **Demerits: -**

a) It cannot be determined by inspection nor it can be located graphically.
b) Arithmetic mean cannot be used if we are dealing with qualitative characteristic which cannot be measured quantitively; such as, intelligence, honesty, beauty, etc. In such cases median (discussed later) is the only average to be used.
c) Arithmetic mean cannot be obtained if a single observation is missing or lost or is illegible unless we drop it out and compute the arithmetic mean of the remaining values.
d) Arithmetic mean is affected very much by extreme values.
e) In extremely asymmetrical (skewed) distribution, usually arithmetic mean is not a suitable measure of location.
f) It cannot be calculated for frequency distribution with open end classes.

FROMULA

No.	Ungrouped data	Grouped data (Discrete & Continuous Data)
1.	$\bar{x} = \frac{\sum x}{n}$	$\bar{x} = \frac{\sum fx}{\sum f}$
2.	$\bar{x} = A + \frac{\sum d}{n}$ where, d = X – A A = Assume no. in X	$\bar{x} = A + \frac{\sum fd}{\sum f}$ where, d = X – A A = Assume no. in X **OR** $\bar{x} = A + \frac{\sum fd}{\sum f} \times c$

EXAMPLE

UNGROUPED DATA

1. The marks obtained by ten students, in an examination are given below. Find mean of the marks. 18, 16, 20, 8, 22, 19, 17, 10, 6, 12

X	18	16	20	8	22	19	17	10	6	12	**148**

$$\bar{x} = \frac{\sum x}{n}$$
$$\bar{x} = \frac{148}{10}$$
$$\bar{x} = 14.8$$

X	15	20	23	24	29	42	55	68	79	58	Total
d = X-A, A = 29	-14	-9	-6	-5	0	13	26	39	50	29	123

2. Find the mean of the following observation.
   15, 20, 23, 24,29, 42, 55, 68, 73, 58.

$$\bar{x} = A + \frac{\sum d}{n}$$
$$\bar{x} = 29 + \frac{123}{10}$$
$$\bar{x} = 29 + 12.3$$
$$\bar{x} = 41.3$$

3. Find the mean of the following observation.
   47, 53, 52, 59, 72, 83, 92, 94, 98, 99 **[Ans : $\bar{x}$ =74.9]**
4. Find the mean of the following observation.
   126, 110, 91, 115, 112, 80, 101, 93, 97, 113 **[Ans : $\bar{x}$ =103.8]**

GROUPED DATA (Discrete Data)

1. The following is the distribution of the number of children in 60 families find the average number of children per family.

Number of children	0	1	2	3	4	5	6
Number of Families	3	8	12	21	10	4	2

Number of children x	Number of families f	fx
0	3	0
1	8	8
2	12	24
3	21	63
4	10	40
5	4	20
6	2	12
Total	**60**	**167**

$$\bar{x} = \frac{\sum fx}{\sum f}$$
$$\bar{x} = \frac{167}{60}$$
$$\bar{x} = 2.78$$

2. Find mean.

Observation	50	51	52	53	54	55	56	57	58
Frequency	1	4	8	13	9	7	4	3	1

observation x	Frequency f	d = X -A A = 54	fd
50	1	-4	-4
51	4	-3	-12
52	8	-2	-16
53	13	-1	-13
54	9	0	0
55	7	1	7
56	4	2	8
57	3	3	9
58	1	4	4
Total	**50**		**-17**

$$\bar{x} = A + \frac{\sum fd}{\sum f}$$
$$\bar{x} = 54 + \frac{-17}{50}$$
$$\bar{x} = 54 + (-0.34)$$
$$\bar{x} = 54 - 0.34$$
$$\bar{x} = 53.66$$

3. Find mean.

Observation	69	89	109	129	149	169	189
Frequency	5	8	9	14	3	1	40

Price x	No. of shops f	$d = \frac{X - A}{c}$, A = 129, c= 20	fd
69	5	-3	-15
89	8	-2	-16
109	9	-1	-9
129	14	0	0
149	3	1	3
169	1	2	2
189	40	3	120
Total	**75**		**100**

$$\bar{x} = A + \frac{\sum fd}{\sum f} \times c$$
$$\bar{x} = 129 + \frac{100}{75} \times 20$$
$$\bar{x} = 129 + 1.33 \times 20$$
$$\bar{x} = 129 + 26.6$$
$$\bar{x} = 155.6$$

4. The price of a item change from shop to shop. The following data are available find the mean price

Price	206	212	218	220	224	230
No. of shops	5	8	9	14	3	1

[Ans : $\bar{x}$ =216.75]

5. The following is the distribution is obtained 60 students od standard 11. find the mean for the distribution.

Age	15	16	17	18	19	20	21
Number of Students	3	10	15	20	8	3	1

[Ans : $\bar{x}$ =17.55]

6. Find mean

Observation	20	19	18	17	16	15	14	13
Frequency	3	10	21	25	19	15	5	2

[Ans : $\bar{x}$ =16.78]

GROUPED DATA (Continuous Data)

1. Calculate the mean of the following frequency distribution.

Class-interval	0-8	8-16	16-24	24-32	32-40	40-48
Frequency	8	7	16	24	15	7

Class-interval	Frequency f	x	d = X - A, A = 28	fd
0-8	8	4	-1.2	-9.6
8-16	7	12	-0.8	-5.6
16-24	16	20	-0.4	-6.4
24-32	24	28	0	0
32-40	15	36	0.4	6
40-48	7	44	0.8	5.6
Total	**77**			**-10**

$$\bar{x} = A + \frac{\sum fd}{\sum f}$$
$$\bar{x} = 28 + \frac{(-10)}{77}$$
$$\bar{x} = 28 + (-0.13)$$
$$\bar{x} = 27.87$$

2. The frequency distribution of the marks in statistics of 100 students is given below. Obtain the average marks from it.

Marsk	5-14	15-24	25-34	35-44	45-54	55-64	65-74
No. of students	8	11	15	25	20	13	8

Marks	No of students f	x	$d = \frac{X - A}{c}$, A = 39.5, c= 10	fd
5-14	8	9.5	-3	-24
15-24	11	19.5	-2	-22
25-34	15	29.5	-1	-15
35-44	25	39.5	0	0
45-54	20	49.5	1	20
55-64	13	59.5	2	26

65-74	8	69.5	3	24
Total	**100**			**9**

$$\bar{x} = A + \frac{\sum fd}{\sum f} \times c$$

$$\bar{x} = 39.5 + \frac{9}{100} \times 10$$

$$\bar{x} = 39.5 + 0.09 \times 10$$

$$\bar{x} = 39.5 + 0.9$$

$$\bar{x} = 40.4$$

3. Obtain the mean of the following frequency distribution.

Class	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	1	7	15	27	20	8	2

[Ans : $\bar{x}$ =36.25]

4. Calculate the mean of the following frequency distribution

Class	0-2	2-5	5-10	10-20	20-50	50-100	100-200	200-500	500-1000
Frequency	2	6	8	20	45	22	18	6	3

[Ans : $\bar{x}$ =81.98]

5. The data about weight of mangoes from a tree are as follows. Find the mean of weight mangoes.

Weight of mangoes	410-420	420-430	430-440	440-450	450-460	460-470
No. of mangoes	17	13	15	20	30	33

[Ans : $\bar{x}$ =441.72]

6. Find the mean of the following frequency distribution

Class	2-6	6-10	10-14	14-18	18-22	22-26
Frequency	6	10	20	15	7	2

[Ans : $\bar{x}$ =12.87]

MISSING FREQUENCY

1. If the mean of the following frequency distribution is 122.7 then find the missing frequency.

Class	60-79	80-99	100-119	120-139	140-159	160-179	180-199
Frequency	7	4	?	18	5	5	2

Class	Frequency f	x	$d = \frac{X - A}{c}$, A = 109.5, c= 20	fd
60-79	7	69.5	-2	-14
80-99	4	89.5	-1	-4
100-119	F1	109.5	0	0
120-139	18	129.5	1	18
140-159	5	149.5	2	10
160-179	5	169.5	3	15
180-199	2	189.5	4	8
Total	**41 + F1**			**33**

Mean value $\bar{x}$ **= 122.7 is given**

$$\bar{x} = A + \frac{\sum fd}{\sum f} \times c$$
$$122.7 = 109.5 + \frac{33}{41 + F1} \times 20$$
$$122.7 - 109.5 = \frac{660}{41 + F1}$$
$$13.2 = \frac{660}{41 + F1}$$
$$13.2(41 + F1) = 660$$
$$541.2 + 13.2F1 = 660$$
$$13.2F1 = 660 - 541.2$$
$$13.2F1 = 118.8$$
$$F1 = \frac{118.8}{13.2}$$
$$F1 = 9$$

So the frequency of class 110 – 119 is 9.

2. If the mean of the frequency distribution is 27.75 then find the missing frequency.

Class	0-10	10-20	20-30	30-40	40-50
Frequency	4	6	?	17	4

ii) Geometric mean

Geometric mean of a set of n observations is the nth root of their product. It is denoted by G

Uses: -

a) Suppose we study a variable that change over time. If you want to find average rate change the variable, the application of the arithmetic mean is not appropriate that time we use the geometric mean.
b) To find the rate of population growth and the rate of interest.
c) It is appropriate for averaging the ratio of the change, for average of the proportion, etc.
d) It is most suitable average for index number.

- **Merits: -**

a) It is rigidly defined.
b) It is based upon all the observation.
c) It is suitable for further mathematical treatment.
d) It is not affected much by fluctuations of sampling.
e) It gives comparatively more weight to small items.
f) It is least affected by extreme values

- **Demerits: -**

a) If anyone of the observations is zero, geometric mean becomes zero.
b) If any one or more values are negative, either geometric mean will not be calculated or absurd value will be obtained.
c) Its calculation in somewhat complicated.

FROMULA

$$G = \sqrt[n]{x_1 \times x_2 \times x_3 \times \dots \dots \times x_n}$$

EXAMPLE

1. The population of an area increased by 15% ,18%, 13% and 20 % respectively for four years. Find the average rate increase in the population. **[Ans. : G = 16.47%]**
2. The value of machine depreciates at the rate of 10%, 7%, 5% and 2% in its first four years respectively. Find the average rate of depreciation using an appropriate method. **[Ans. : G = 6.05%]**
3. Find the geometric mean from the following observation.
   12,18,32,36 **[Ans. : G = 22.33]**

iii) Harmonic mean

Harmonic mean of number of observations is the reciprocal of the arithmetic mean of the reciprocals of the given values.

Harmonic mean is the inverse of the arithmetic mean of the reciprocals of the observations of set.

Uses: -

Harmonic mean is suitable when the values are pertaining to the rate of change per unit time such as speed, number of items produced per day, contracts completed per year, etc. In general, harmonic mean is suitable for time, speed, rates, prices, etc.

- **Merits: -**

a) It is based on all the observations of a set.
b) It is a good mean for a highly variable series.
c) It gives more weightage to the small values and less weightage to the large values.
d) It is better than weighted mean since in this value are automatically weighted.
e) It is suitable for further mathematical treatment.
f) It is useful only when small items have to be given a greater weightage.

- **Demerits: -**

a) Harmonic mean it is not easily understood and is difficult to compute.
b) If the any value is zero, it cannot be calculated.

FROMULA

No.	Ungrouped data	Grouped data (Discrete & Continuous Data)
1.	$H.M = \frac{n}{\sum\left(\frac{1}{x}\right)}$	$H.M = \frac{\sum f}{\sum\left(\frac{f}{x}\right)}$

EXAMPLE

UNGROUPED DATA

1. Find the harmonic mean from the following observation.
   2574, 475, 75, 5, 0.8, 0.08, 005, 0.0009 **[Ans. : H.M. = 0.006]**
2. Calculated the harmonic mean from the following observation.

3834, 382, 63, 8, 0.4, 0.03, 0.009,0.0005 **[Ans. : H.M. = 0.0037]**

GROUPED DATA (Discrete Data)

1. From the following data computer value of harmonic mean.

Marks	10	20	25	40	50
Number of students	20	30	50	15	5

[Ans. : H.M. = 20.08]

2. From the following data obtained from the survey. computer harmonic mean.

Speed of the car	130	135	140	145	150
Number of cars	20	30	50	15	5

[Ans. : H.M. =]

GROUPED DATA (Continuous Data)

1. From the following data computer value of harmonic mean.

Class	10-20	20-30	30-40	40-50	50-60
Frequency	4	6	10	7	3

[Ans. : H.M. =]

2. From the following data computer value of harmonic mean.

Class	2-6	6-10	10-14
Frequency	10	12	18

[Ans. : H.M. =]

iv) Median

Median is defined as the value of Middle east observation when the data are arranged in the either ascending or descending order. It is most preferable measure of the location for asymmetric distribution.

- **Merits: -**

a) Median is the only average to be used while dealing with qualitative data.
b) It is rigidly defined.
c) It is easily understood and easily calculated.
d) It is not at all affected by its value.
e) It can be calculated for a distribution with open index class.
f) Median can be located even if the data are uncompleted.

- **Demerits: -**

a) In case of the even number of observation median cannot be determined exactly We merely estimated it by taking the mean of the two middle terms.
b) It is not based on the all observations.
c) It is not used for further algebraic treatment.
d) As compared with mean, it is affected much by fluctuation of the sampling.

FROMULA

No.	Ungrouped data	Grouped data (**Discrete**)
1.	$M = \left(\frac{n+1}{2}\right)^{th} observation$	$M = \left(\frac{n+1}{2}\right)^{th} observation$ $n = \sum f$
		Grouped data (**Continuous)** $class\ of\ M = \left(\frac{n}{2}\right)^{th} observation$ $M = L + \frac{\left(\frac{n}{2}\right) - cf}{f} \times c$ Where, L = lower boundary point of Midian class Cf = cumulative frequency of the class prior to . median class

		F = frequency if median class C = length of median class

EXAMPLE

UNGROUPED DATA

1. The weight of 9 Pearsons are as follows. Find median weight.
   52, 48, 60, 50, 68, 51, 55, 62, 58
   Ans: - Arrangement of observation in ascending form is as follows.
   48, 50, 51, 52, 55, 58, 60, 62, 68.

$$\text{Median } M = \left(\frac{n+1}{2}\right)^{th} observation$$
$$M = \left(\frac{9+1}{2}\right)^{th} observation$$
$$M = (5)^{th} observation$$
$$M = 55\ kg$$

2. The marks obtained by 12 students, in an examination are given below. Find median of the marks. 4, 6, 5, 8, 12, 7, 10, 5, 15, 9, 10, 11
   Ans: - Arrangement of observation in ascending form is as follows.
   4, 5, 5, 6, 7, 8, 9, 10, 10, 11, 12, 15.

$$\text{Median } M = \left(\frac{n+1}{2}\right)^{th} observation$$
$$M = \left(\frac{12+1}{2}\right)^{th} observation$$
$$M = (6.5)^{th} observation$$
$$M = \frac{6^{th}\ obs. + 7^{th}\ obs.}{2}$$
$$M = \frac{8\ + 9}{2}$$
$$M = 8.5\ marks$$

3. The number of items produced by factory in different weeks are 85, 92, 68, 80, 72, 63, and 55 find the median of the production. **[Ans. : M = 80]**
4. Find the median of the following observation.
   13, 17, 22, 8, 19, 14, 20, 6 **[Ans. : M = 15.5]**
   Find median. 8, 6, 7, 0, 2, 4, 6, 5, 5, 4, 8, 9, 3, 6, 7. **[Ans. : M = 6]**

5.

GROUPED DATA (Discrete Data)

1. The frequency distribution of the number of mistakes committed by 335 candidates in typing examination is given below find the median of the distribution

Number of mistakes	0	1	2	3	4	5	6	7	8
No. of candidates	35	82	107	40	31	15	12	9	4

No. of mistakes	No. of candidates	CF
0	35	35
1	82	117
2	107	224
3	40	264
4	31	295
5	15	310
6	12	322
7	9	331
8	4	335
Total	**335**	

$$M = \left(\frac{n+1}{2}\right)^{th} observation$$
$$M = \left(\frac{335+1}{2}\right)^{th} observation$$
$$M = (168)^{th} observation$$

From the column of CF it is seen that the value of observation numbers from 118 to 224 is 2. hence the value of 168 observation is also 2.

M = 2 mistakes

2. The time require for typing or report by the different typists is given in the following data. Find the median typing time using it.

Time for typing	10	11	12	13	14
No. of typists	5	7	8	15	5

Time for typing	No. of typists	CF
10	5	5
11	7	12
12	8	20
13	15	35
14	5	40
Total	**40**	

$$M = \left(\frac{n+1}{2}\right)^{th} observation$$

$$M = \left(\frac{40+1}{2}\right)^{th} observation$$

$$M = (20.5)^{th} observation$$

$$M = \frac{20^{th}\ obs. + 21^{th}\ obs.}{2}$$

$$M = \frac{12\ + 13}{2}$$

$$M = 12.5\ marks$$

3. We have the following information from the survey of hundred customers from the bank their number of visits to bank during the month

No. of visits	0	1	2	3	4	5	6
No. of customer	12	22	40	15	6	4	1

No. of visits	No. of customer	CF
0	12	12
1	22	34
2	40	74
3	15	89
4	6	95
5	4	99
6	1	100
Total	**100**	

$$M = \left(\frac{n+1}{2}\right)^{th} observation$$

$$M = \left(\frac{100+1}{2}\right)^{th} observation$$

$$M = (50.5)^{th} observation$$
$$M = 2$$

4. The following table shows the record of the absent students of the class during a month. Find the median of the number of absent days per students

No. of absent days of students	0	1	2	3	4	5
No. of students	8	12	18	9	5	1

[Ans. : M = 2 days]

5. The following table give ages 80 students selected from a college.

Age	17	18	19	20	21	22	23
No. of Students	11	14	22	15	8	6	12

[Ans. : M = 19 students]

GROUPED DATA (Continuous Data)

1. Calculate the median of the following frequency distribution.

Class-interval	10-29	30-49	50-69	70-89	90-109	110-129
Frequency	13	22	48	57	25	5

Class-interval	Frequency	CF
10-29	13	13
30-49	22	35
50-69	48	83
70-89	57	140
90-109	25	165
110-129	5	170
Total	**170**	

$$class\ of\ M = \left(\frac{n}{2}\right)^{th} observation$$
$$= \left(\frac{170}{2}\right)^{th} observation$$
$$= (85)^{th} observation$$

From the cf column 85th observation lies in the class 70-89.

Taking limit points median class = 70 – 89

= 69.5 – 89.5

$$M = L + \frac{\left(\frac{n}{2}\right) - cf}{f} \times c$$
$$M = 69.5 + \frac{85 - 83}{57} \times 20$$
$$M = 69.5 + \frac{40}{57}$$
$$M = 69.5 + 0.70$$
$$M = 70.20$$

2. The data about monthly expenditure on petrol for 75 families is given in the following table. Find the medial expenditure on petrol for these families.

Expenditure on petrol	Upto 200	Upto 400	Upto 600	Upto 800	Upto 1000	Upto 1200
No. of families	2	8	17	32	57	75

Expenditure on petrol	No. of families	CF
upto 200	2	2
200-400	6	8
400-600	9	17
600-800	15	32
800-1000	25	57
1000-1200	18	75
Total	**75**	

$$class\ of\ M = \left(\frac{n}{2}\right)^{th} observation$$
$$= \left(\frac{75}{2}\right)^{th} observation$$
$$= (37.5)^{th} observation$$

From the cf column 37.5th observation lies in the class 800-1000.

Taking limit points median class = 800 - 1000

$$M = L + \frac{\left(\frac{n}{2}\right) - cf}{f} \times c$$
$$M = 800 + \frac{37.5 - 32}{25} \times 200$$
$$M = 800 + \frac{1100}{25}$$
$$M = 800 + 44$$
$$M = 844$$

3. The level of air pollution in a city on different days is as follows find the median level of pollution.

Level of pollution	250 and above	270 and above	290 and above	310 and above	320 and above	330 and above	340 and above
No. of days	150	133	108	76	41	20	7

Level of pollution	No. of days	CF
250-270	17	17
270-290	25	42
290-310	32	74
310-320	35	109
320-330	21	130
330-340	13	143
340 & above	7	150
Total	**150**	

$$class\ of\ M = \left(\frac{n}{2}\right)^{th} observation$$
$$= \left(\frac{150}{2}\right)^{th} observation$$
$$= (75)^{th} observation$$

From the cf column 75th observation lies in the class 310-320.

Taking limit points median class = 310-320

$$M = L + \frac{\left(\frac{n}{2}\right) - cf}{f} \times c$$

$$M = 310 + \frac{75 - 74}{35} \times 10$$
$$M = 310 + \frac{10}{35}$$
$$M = 310 + 0.2857$$
$$M = 310.29$$

4. Find median from the following frequency distribution.

class	Less than 20	20-40	40-60	60-80	80-100	More than 100
No. of students	8	11	15	25	20	13

[Ans. : M = 53.33]

5. Find median age.

Age	55-60	50-55	45-50	40-45	35-40	30-35	25-30	20-25
No. of Pearsons	17	13	15	20	30	33	28	14

[Ans. : M = 35.83 year]

6. Find median of the following distribution.

Class	118-126	127-135	136-144	145-153	154-162	163-171	172-180
Frequency	17	13	15	20	30	33	28

[Ans. : M = 146.75]

7. Use the following guitar to find a median salary of employees in a firm.

salary	5 or more	10 or more	15 or more	20 or more	25 or more	30 or more
No. of employees	150	133	108	76	41	20

[Ans. : M = 21.78]

8. The following table shows the monthly expense for entire any of group of hundred students find the median of these expense

Expense	Less than 200	Less than 400	Less than 600	Less than 700	Less than 800	Less than 900
No. of students	8	31	71	88	95	100

[Ans. : M = 495]

MISSING FREQUENCY

1. Following is an incomplete frequency distribution marks obtained by the 120 students in any one examination if the median score of the student is 75 then find the missing frequency

Marks	30-39	40-49	50-59	60-69	70-79	80-89	90-99
Students	1	3	?	20	?	33	9

Marks	Students	CF
30-39	1	1
40-49	3	4
50-59	F1	4 + F1
60-69	20	24 + F1
70-79	F2	24 + F1+ F2
80-89	33	57 + F1 + F2
90-99	9	66 + F1 + F2
Total	**120**	

Here, n = 120

66 + F1 + F2 = 120
F1 + F2 = 120 – 60
F1 + F2 = 54
F1 = 54 - F2 ……. eq. (1)

M = 75 is given
class of M = 70 - 79
= 69.5 – 79.5

$$M = L + \frac{\left(\frac{n}{2}\right) - cf}{f} \times c$$

$$75 = 69.5 + \frac{60 - (24 + f1)}{f2} \times 10$$

$$75 - 69.5 = \frac{60 - 24 - f1}{f2} \times 10$$

$$5.5 = \frac{36 - f1}{f2} \times 10$$

$$5.5f2 = (36 - f1) \times 10$$

$$5.5f2 = 360 - 10f1 \qquad \ldots\ldots eq(2)$$

Now, from equation (1) f1 = 54 – f2 put in the equation (2)

$$5.5f2 = 360 - 10(54 - f2)$$
$$5.5f2 = 360 - 540 + 10f2$$
$$5.5f2 - 10f2 = 360 - 540$$
$$5.5f2 - 10f2 = 360 - 540$$
$$-4.5f2 = -180$$
$$f2 = \frac{-180}{-4.5}$$
$$f2 = 40$$

Now $f2 = 40$ value put in equation (1)

$$f1 = 54 - f2$$

$$f1 = 54 - 40$$
$$f1 = 14$$

2. If the median of the frequency distribution is 146.75 then find the missing frequency.

Length	118-126	127-135	136-144	145-153	154-162	163-171	172-180	Total
Frequency	10	20	?	40	?	25	16	40

[Ans. : f1 = 12 & f2 = 5]

3. If the median of the frequency distribution is 35 and total frequency = 170 then find the missing frequency.

salary	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of workers	10	20	?	40	?	25	16

[Ans. : f1 = 35 & f2 = 24]

v) <u>Mode</u>

The value which data repeated maximum number of I term both the value occurring with maximum frequency in a given data is called a mode.

- **Merits: -**

a) Mode is readily comprehensible and easy to calculate. Like median, mode can be located in some cases merely by inspection.
b) Mode is not at all affected by extreme values.
c) Mode can be conveniently located even if the frequency distribution has class-intervals of unequal magnitude provided the modal class and the classes preceding

and succeeding it are of the same magnitude. Open end classes also do not pose any problem in the location of mode.

- **Demerits: -**

a) Mode is ill-defined. It is not always possible to find a clearly defined mode. In Some cases, we may come across distributions with two modes. Such distributions are called bi-modal. If a distribution has more than two modes, it is said to be multimodal.
b) It is not based upon all the observations.
c) It is not capable of further mathematical treatment.
d) As compared with mean, mode is affected to a greater extent by fluctuations of sampling.

FROMULA

UNGROUPED DATA

M_0 = The observation which is required for maximum number of times

GROUPED DATA(DISCREATE)

M_0 = maximum frequency of the distribution

GROUPED DATA(CONTINUOUS)

$$M_0 = L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times c$$

Where,

$L \quad = Lower\ limit\ of\ the\ modal\ class$
$f_m = Frequency\ of\ modal\ class$
$f_1 = \ Frequency\ of\ the\ class\ preceding\ the\ model\ class$
$f_2 = Frequency\ of\ the\ class\ succeeding\ the\ model\ class$
$c \ = class\ lengh\ of\ the\ modal\ class$

- **The above formula we can not used when following condition appear in the data**
  a) More than one observation in a frequency distribution appears with highest frequency
  b) The continuous distribution has classes of unequal length.
  c) The frequency distribution is mixed distribution that is a part of it is discrete and the rest is continuous

d) The right or left end of the curve of the frequency distribution is too extended.

We have observed that mode is not well, defined in many cases. The noted statistician Karl Pearson established a relation between mean, median and mode by studying their vales for different data sets, He observed that for data that are not evenly distributed around average, difference between mean and mode is approximately 3 times the difference between mean and median.

That is, (Mean - Mode) =3 (Mean - Median)

The following formula is obtained to find the mode using these relations

Mode - 3 (Median) - 2(Mean)

This is written in notations as $M_0 = 3M - 2\bar{x}$

This formula to find mode is called as an empirical formula because the value obtained from observation and not from the theory. The value of mode found using this formula can be negative if the frequency distribution is not evenly distributed around the average.

EXAMPLE

UNGROUPED DATA

1. The number of accidents occur during last 15 days in a city are given below. obtain mode of the series
   2, 3, 0, 1, 4, 0, 1, 3, 1, 2, 5, 4, 1, 0, 1
   Ans: - Here observation 1 is repeated for maximum number of times
   Mode Mo = 1
2. The following figures are the number of mistakes committed by typist in different pages. find mode of the mistake.
   2, 1, 3, 3, 4, 2, 0, 3, 2, 1, 2, 0, 3, 5, 4 **[Ans. : M_0 = 1 & 2]**
3. The number of books purchased by each of the 15 persons from the bookstore are followers
   1, 0, 2, 2, 3, 4, 2, 7, 2, 2, 5, 4, 2, 1, 2. **[Ans. : M_0 = 2]**

GROUPED DATA (Discrete Data)

1. Find the mode of the following frequency distribution.

Observation	10	11	12	13	14	15	16
Frequency	15	30	38	52	40	12	10

Ans: - In the given frequency distribution maximum frequency is 52 and it is against observation 13

Mode Mo = 13

2. Find the mode of the following frequency distribution.

Observation	10	11	12	13	14	15	16	17	18
Frequency	3	10	19	28	13	11	7	2	1

[Ans. : M_0 = 13]

3. The Number of trips made by 24 taxis driver in our day are shown in the following data find the mode number of trip.

No. of trips	1	2	4	5	6	7
No. of drivers	3	7	4	7	2	1

[Ans. : M_0 = 2 & 5]

GROUPED DATA (Continuous Data)

1. The following table gives output of workers in a factory find the mode output.

Output	150–160	160-170	170-180	180-190	190-200	200-210	210-220	220-230
No. of workers	4	5	19	33	48	22	12	6

Ans: - The maximum frequency 48 is in the class 190 – 200

L = 190, f_m = 48, f_1 = 33, f_2 = 22, C = 10

$$M_0 = L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times c$$

$$M_0 = 190 + \frac{48 - 33}{2(48) - 33 - 22} \times 10$$

$$M_0 = 190 + \frac{150}{41}$$

$$M_0 = 190 + 3.66$$

$$M_0 = 193.6$$

2. Find mode from the following frequency distribution.

Class	10-14	15-19	20-24	25-29	30-34	35-39
No. of students	12	24	35	15	8	2

[Ans. : M_0 = 21.27]

3. The table below shows the data about the weight of 86 apples from a garden find the mood from the weight of apples.

Weight of apples	120–130	130-140	140-150	150-160	160-170	170-180	180-190
No. of Apples	8	13	19	23	10	8	5

[Ans. : M_0 = 152.35]

4. Find mode from the following frequency distribution.

Class	0-5	5-10	10-20	20-30	30-50	50-70	70-100
Frequency	4	7	12	20	18	12	7

Class	Frequency	x	$d = \frac{X-A}{c}$, A = 25, c= 5	fd	CF
0-5	4	2.5	-4.5	-18	4
5-10	7	7.5	-3.5	-24.5	11
10-20	12	15	-2	-24	23
20-30	20	25	0	0	43
30-50	18	40	3	54	61
50-70	12	60	7	84	73
70-100	7	85	12	84	80
Total	**80**			**155.5**	

Mean

$$\bar{x} = A + \frac{\sum fd}{\sum f} \times c$$

$$\bar{x} = 25 + \frac{155.5}{80} \times 5$$

$$\bar{x} = 25 + 1.94 \times 5$$

$$\bar{x} = 25 + 9.7$$

$$\bar{x} = 34.7$$

Median

$$class\ of\ M = \left(\frac{n}{2}\right)^{th} observation$$

$$= \left(\frac{80}{2}\right)^{th} observation$$

$$= (40)^{th} observation$$

From the cf column 40 observation lies in the class 20 - 30.

Taking limit points median class = 20 - 30

$$M = L + \frac{\left(\frac{n}{2}\right) - cf}{f} \times c$$
$$M = 20 + \frac{40 - 23}{20} \times 10$$
$$M = 20 + \frac{170}{20}$$
$$M = 20 + 8.5$$
$$M = 28.5$$

Mode

$$\boldsymbol{M_0 = 3M - 2\bar{x}}$$
$$\boldsymbol{= 3(28.7) - 2(34.7)}$$
$$\boldsymbol{= 85.5 - 69.44}$$
$$\boldsymbol{= 16.06}$$

5. The time between placing and order and it is delivery was noted for the certain wholesaler as follows find the mode of this time.

Time	20-25	25-30	30-35	35-40	40-45	45-50	50-55
No. of orders	2	5	7	5	6	7	3

[Ans. : $\bar{x}$ = 38.36, M = 38.5, M_0 = 38.78]

6. Find mode from the following frequency distribution.

Age	50-60	60-65	65-70	70-85	85-100
No. of persons	6	10	19	9	4

1 **[Ans. : $\bar{x}$ = 68.85, M = 67.11, M_0 = 63.63]**

MISSING FREQUENCY

1. The incomplete frequency distribution of monthly expenditure detail of 360 families is as follows. If their mode expenditure his rupees 1376 then find the missing frequency.

Expense	Less than 400	400-800	800-1200	1200-1600	1600-2000	2000-2400	2400-2800	2800 and more
No. Of families	14	22	?	124	?	32	15	5

[Ans. : f1 = 80 & f2 = 68]

2. If the mode of the frequency distribution is 341. then find the missing frequency.

Length	200-210	210-220	220-230	230-240	240-250	250-260	260-270	Total
Frequency	4	16	?	?	40	6	4	230

[Ans. : f1 = 60 & f2 = 100]

3. If the mode of the frequency distribution is 74 and total frequency = 163 then find the missing frequency.

Class	40-50	50-60	60-70	70-80	80-90	90-100	100-110	110-120
No. of workers	4	12	?	50	?	13	9	4

[Ans. : f1 = 38 & f2 = 32.6]

MEASURES OF DISPERSION

❖ Meaning of dispersion.

A measure which shows how far the observations of the data are scattered from the measure of average is terms as dispersion.

The degree to which numerical data trend to spread about an average value is called variation or dispersion of the data.

Dispersion or spread is the degree of the scatter or variation of the items.

❖ Characteristics for an Ideal measure of Dispersion

i) It should he rigidly defined.
ii) It should be easy to calculate and easy to understand.
iii) It should be based on all the observations.
iv) It should be amenable to further mathematical treatment.
v) It should be affected as little as possible by fluctuations of sampling.

❖ Concept of Absolute Measure and Relative measure

➢ Absolute Measure: -

A measure of dispersion which is expressed in the same unit in which the observation of the data is expressed is called an absolute measure of dispersion.

Example: - Height in cm, weight in kg, Income in rupees, etc.

In this case two series having different unit of dispersion cannot be compered. Absolute measure is not free from the unit measurement.

➢ Relative measure: -

A measure of dispersion which is free from the unit of measurement is called relative measure of dispersion. It is also known as coefficient of dispersion.

The variability of two or more sets of data having different units of measurement can compared only by relative measure of dispersion.

It is independent of units of measurement of the observation of data.

❖ Measures of dispersion

i) Range
ii) Standard deviation
iii) Variance

Range and quartile deviation are the positional measure of dispersion while mean deviation and standard deviation are known as the measures of dispersion representing the summary of devotions of observations from the measure of central tendency

i) Range

The difference between the largest and smallest values of a set of data is called its range.

Range = largest value - lowest value

Coefficient of range = largest value - lowest value / largest value + lowest value

Lesser the Coefficient of range, better it is.

Range is useful in statistical quality control for the construction of control charts which measure the variability within the samples taken from the production. If variation is not very large then range is useful in measuring variation in money rates, exchange rates, share prices etc. In our day-to-day problem like 'daily sales in a supermarket', 'temperature of a city', 'expense of petrol by two-wheeler or car', etc. are generally expressed in the form of the interval in which it lies, and range of the data can be known from it.

Merits: -

a) It is the easiest measure of dispersion.
b) It can always be found out visually.
c) It is one of the largely used measure of dispersion.

Demerits: -

a) It depends on two extreme values of a series. Thus, it gives no information about the observations lying between smallest and largest values.
b) It is highly susceptible to sampling fluctuations.
c) It is not suitable for further mathematical treatment.
d) It cannot be calculated for the frequency distribution having open-ended classes.

FORMULA

Absolute Range = $x_H - x_L$

Relative Range = $\frac{x_H - x_L}{x_H + x_L}$

EXAMPLE

UNGROUPED DATA

1. The run score by batsmen in his last 10 innings of cricket matches are given below. Find the range and relative range of run.
   48, 75, 37, 52, 93, 81, 25, 72, 18, 60.
   Ans: - $x_H = 93 \quad x_L = 18$

   Absolute Range = $x_H - x_L$

   $$= 93 - 18$$
   $$= 75$$

   Relative Range = $\frac{x_H - x_L}{x_H + x_L}$

   $$= \frac{93 - 18}{93 + 18}$$
   $$= \frac{75}{111}$$
   $$= 0.68$$

2. The following data refers to the height of 10 students of the class. Find the range and relative range of height of students.
   162, 145, 170, 181, 167, 151, 175, 185, 169, 156.

GROUPED DATA (DISCRETE)

1. Find absolute range and relative range of the following data.

Observation	100	101	102	103	104	105	106	107
Frequency	8	21	36	50	28	20	10	5

   Ans: - $x_H = 100 \quad x_L = 107$

   Absolute Range = $x_H - x_L$

   $$= 107 - 100$$
   $$= 7$$

$$\text{Relative Range} = \frac{x_H - x_L}{x_H + x_L}$$

$$= \frac{107 - 100}{107 + 100}$$

$$= \frac{7}{207}$$

$$= 0.0034$$

2. A bus company has 77 busses who travel in the city. The information of the number of passengers in the bus on the particular days at a particular time is given below. Find the range and the relative range of the number of passengers.

No. of passengers	2	7	10	18	25	30	37
No. of buses	1	4	11	17	23	16	5

GROUPED DATA (CONTINUOUS)

1. Find absolute range and relative range of the following data.

Class	44-46	46-48	48-50	50-52	52-54	54-56
Frequency	7	15	35	30	8	2

Ans: - $x_H = 56 \quad x_L = 44$

$$\text{Absolute Range} = x_H - x_L$$

$$= 56 - 44$$

$$= 12$$

$$\text{Relative Range} = \frac{x_H - x_L}{x_H + x_L}$$

$$= \frac{56 - 44}{56 + 44}$$

$$= \frac{12}{100}$$

$$= 0.12$$

2. Find absolute range and relative range of the following data.

Class	10-15	15-20	20-25	25-30	30-35
Frequency	8	15	26	47	4

3. Find the range and the relative range of the marks from the following frequency distribution of the Marks score by 50 students of the school in certain examination.

Marks	50-59	60-69	70-79	80-89	90-99
No. of students	2	15	23	6	4

ii) Standard Deviation (S.D.) & Variance

We have seen that the definition of mean deviation is based on the absolute value of the deviation of observations of the data from the mean. Since the algebraic sign of the observation are ignored, mean deviation is less used in advanced study of statistics. This limitation of Mean deviation is overcome by an important measure of dispersion known as the Standard Deviation. Instead of talking the absolute value of deviation each observation from the mean the square of the deviation is taken. If the sum of the square of this deviation is divided by the total number of observations, we get an important measure of dispersion known as the variance. This positive square root of the variance is called as the standard deviation.

Well known statistician Karl Pearson defined the "standard deviation as the standard deviation is the positive square root of the mean of the square of the deviation's measures from the mean."

After mean, standard deviation is another very useful measure which gives information about values of the observation of a population.

Standard deviation is also known as mean error, mean square error or root mean square deviation from mean. As a matter of fact, S.D. is nothing but the quadratic mean of the deviation from the data.

Merits: -

a) All the observation are used in its computation.
b) Standard deviation is the most effective and widely used measure of dispersion among all the measures of dispersion.
c) Standard deviation is a suitable measure for algebraic calculation.
d) Standard deviation carries great importance in sampling methods.
e) It is least sensitive to sampling fluctuations.
f) With the help of S.D., it is possible to ascertain the area under the normal curve.
g) It has great utility in testing of hypotheses which other measures of dispersion hardly do.

Demerits: -

a) The extreme observations get undue importance in the value this measure.
b) It cannot be obtained if the frequency distributions have open-ended classes.

FROMULA

No.	Ungrouped data	Grouped data (Discrete & Continuous Data)
	Standard Deviation Formula	
1.	$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$	$s = \sqrt{\frac{\sum f(x-\bar{x})^2}{n}}$
2.	$s = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$	$s = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$
3.	$s = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$	$s = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$ $\boldsymbol{whene\ d = x - A}$ $s = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2} \times C$ $\boldsymbol{whene\ d = \frac{x-A}{c}}$
	Variance Formula	
1.	$s^2 = \frac{\sum(x-\bar{x})^2}{n}$	$s^2 = \frac{\sum f(x-\bar{x})^2}{n}$
2.	$s^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$	$s^2 = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$
3.	$s^2 = \frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2$	$s^2 = \frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2$ $\boldsymbol{whene\ d = x - A}$ $s^2 = \frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2 \times C$ $\boldsymbol{whene\ d = \frac{x-A}{c}}$

EXAMPLE

UNGROUPED DATA

1. The run score by a batsman in his last seven matches are given below 52, 58, 40, 60, 54, 38, 48. find the variance and the runs of batsmen and also find the standard deviation

X	$x - \bar{x}$	$(x - \bar{x})^2$
52	2	4
58	8	64
40	-10	100
60	10	100
54	4	16
38	-12	144
48	-2	4
350		**432**

Mean $\bar{x} = \frac{\sum x}{n}$

$$= \frac{350}{7}$$

$$= 50\ runs$$

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

$$= \sqrt{\frac{432}{7}}$$

$$= \sqrt{61.41}$$

$$= 7.86\ runs$$

Variance $s^2 = (7.86)^2$

$$= 61.78\ runs$$

2. The time taken to slow puzzle by the 5 students are 5, 8, 3, 6, 10. Compute standard deviation of the time taken to solve the puzzles and also find the variance.

X	X^2
5	25
8	64
3	9
6	36

10	100
32	**234**

Mean $\bar{x} = \frac{\sum x}{n}$

$= \frac{32}{5}$

$= 6.4\ minutes$

Standard deviation $s = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$

$= \sqrt{\frac{234}{5} - \left(\frac{32}{5}\right)}$

$= \sqrt{46.8 - 40.96}$

$= \sqrt{5.84}$

$= 2.42\ minutes$

Variance $s^2 = (2.42)^2$

$= 5.86\ minutes$

3. The following are closing prices of 5 shares 132, 147 120, 152 and 125. Find standard deviation and variance.

X	$d = x - A$ A = 120	d^2
132	12	144
147	27	729
120	0	0
152	32	1024
125	5	25
676	**76**	**1922**

Standard deviation $s = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$

$= \sqrt{\frac{1922}{5} - \left(\frac{76}{5}\right)}$

$= \sqrt{384.4 - 231.04}$

$= \sqrt{153.36}$

$= 12.38$

Variance $s^2 = (12.38)^2$

$= 153.26$

4. The monthly pocket expense of 10 students are rupees 225, 190, 220, 245, 270, 180, 250, 300, 175 and 265 find the standard deviation of the expense and also find the variance of the expense. **[Ans: s = 35.45 & s^2 = 1556]**
5. Find the S.D. and variance of the following data.
   60, 45, 25, 40, 70, 32. **[Ans: s = 15.51 & s^2 = 240.55]**
6. A time of the 10 competitors in 400 meters running rays are respectively 70, 75, 90, 95, 72, 80, 88, 85, 75 and 80 seconds. Find the variance and standard deviation of the data. **[Ans: s = 7.86 & s^2 = 61.78]**

GROUPED DATA (DISCRETE DATA)

1. The distribution of the number of absent days of 15 students of a class in the month of January is given below find a standard deviation and variance of their numbers of absent students.

No. of absent days	0	1	2	3	4
No. of students	1	3	7	3	1

No. of absent days x	No. of students f	fx	$x - \bar{x}$	$(x - \bar{x})^2$	$f(x - \bar{x})^2$
0	1	0	-2	4	4
1	3	3	-1	1	3
2	7	14	0	0	0
3	3	9	1	1	3
4	1	4	2	4	4
Total	**15**	**30**	**0**	**10**	**14**

Mean $\bar{x} = \frac{\sum fx}{\sum f}$

$= \frac{30}{15}$

$= 2\ days$

Standard deviation $s = \sqrt{\frac{\sum f(x-\bar{x})^2}{\sum f}}$

$= \sqrt{\frac{14}{15}}$

$$= \sqrt{0.9333}$$
$$= 0.97\ days$$

Variance $s^2 = (0.97)^2$
$$= 0.94\ days$$

2. The following frequency distribution represent the amount of deposit and the number of depositors in the bank find the standard deviation and variance of the deposits.

Deposits	5	10	15	20	25	30	35
No of depositors	2	7	11	15	10	4	1

Observation x	Frequency f	fx	Fx^2
5	2	10	50
10	7	70	700
15	11	165	2475
20	15	300	6000
25	10	250	6250
30	4	120	3600
35	1	35	1225
Total	**50**	**950**	**20300**

Standard deviation $s = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$
$$= \sqrt{\frac{20300}{50} - \left(\frac{950}{50}\right)^2}$$
$$= \sqrt{406 - 361}$$
$$= \sqrt{45}$$
$$= 6.71$$

Variance $s^2 = (6.71)^2$
$$= 45.02$$

Using other formula

Observation x	Frequency f	$d = x - A$ A = 20	fd	fd^2
5	2	-15	-30	450
10	7	-10	-70	700
15	11	-5	-55	275
20	15	0	0	0
25	10	5	50	250
30	4	10	40	400
35	1	15	15	225
Total	50		-50	2300

Standard deviation $s = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$

$$= \sqrt{\frac{2300}{50} - \left(\frac{-50}{50}\right)^2}$$

$$= \sqrt{46 - 1}$$

$$= \sqrt{45}$$

$$= 6.71$$

Variance $s^2 = (6.71)^2$

$$= 45.02$$

3. Find the standard deviation.

Observation	62	67	72	77	82	87	92
Frequency	5	7	10	8	5	3	2

[Ans: s = 8.135]

4. The information of the number of mobile phones sold in 35 days in the mobile shop is given below find the standard deviation variance of number of mobile phones sold.

No. of mobile sold	5	6	7	8	9	10
No. of days	2	5	8	12	7	1

[Ans: s = 1.20 & s^2 = 1.45]

5. Find S.D. and variance.

Observation	70	60	50	40	30	20	10
Frequency	7	18	40	45	63	68	75

[Ans: s = 17.39 & s^2 = 302.41]

GROUPED DATA (CONTINUOUS DATA)

1. Find standard deviation from the following frequency distribution of marks obtained by 200 students of a school in an examination.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of students	5	12	30	45	50	37	21

Marks	No, of students f	X	$d = \frac{x - A}{c}$ A = 35, c =10	fd	fd^2
0-10	5	5	-3	-15	45
10-20	12	15	-2	-24	48
20-30	30	25	-1	-30	30
30-40	45	35	0	0	0
40-50	50	45	1	50	50
50-60	37	55	2	74	148
60-70	21	65	3	63	189
Total	**200**			**118**	**510**

Standard deviation $s = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2} \times C$

$$= \sqrt{\frac{510}{200} - \left(\frac{118}{200}\right)^2} \times 10$$

$$= \sqrt{2.55 - 0.3481} \times 10$$

$$= \sqrt{2.2019} \times 10$$

$$= 14.84 \text{ marks}$$

2. Find the standard deviation of the daily wages from following information of wages workers of a factory.

Daily wages	130-150	150-170	170-190	190-120	210-230
Frequency	7	18	40	45	63

3. Find the standard deviation following frequency distribution.

Class	100-200	200-300	300-400	400-500	500-600	600-700	700-800
Frequency	13	25	38	45	36	30	13

[Ans: s = 162.43]

4. Find the standard deviation of the age of the persons from the following frequency of 125 persons living in the society.

Age	0-10	10-20	20-0	30-40	40-50	50-60	60-70	70-80
No. of persons	15	15	23	22	25	10	5	10

[Ans: s =19.76]