2.1 STATISTICS : ITS INTRODUCTION & SOME MEASURES (Ground work for subsequent chapters)

2.1.1. Statistics and its Related Terms

Statistics. An aggregate of facts which are affecetd by a number of causes and which are expressed numerically to some resonable extent of accuracy and which are collected in a systematic manner for a specific purpose is called statistics.

The subject in which we study the characteristics of the facts is also known as Statistics .

Illustration. Let we collect the facts and figures of the car accident taken place in Kolkata during last 15 years. Then this collection is a *statistics of 'car accident in Kolkata during last 15 years'.*

Variable. Variable is a symbol *e.g.* x, y...that can assume any prescribed value.

If a variable assume only one value then it is called *constant.*

Illustration. (i) Let N = number of members in a family of India. Then N can assume any of the values 1, 2, 3, ... So, here N is a variable.

(ii) Let X = number of prime ministers in India .Then we see X assumes only one value,1.So X is a constant.

(iii) Let H = weight of a person of city. Then we see H can assume the values 45 kg, 46.12 kg, 80.0015 kg etc.So here H is a variable.

Discrete and Continuous Variable. A variable that can theoretically assume any value between two given values is called a continuous variable.

A variable which is not continuous is called a discrete variable.

In the above illustration the variable cited in (i) and (ii) are discrete whereas the variable in (iii) is continuous.

Data / Observations. The values assumed by a variable are known as data or observation. Sometimes, in statistics, these are regarded as statistical data or statistical observation.

Illustration. let Y = marks obtained by the students in Mathematics in a class. Then Y is a variable which can assume the datas 5,90,0,81 etc .The data can be presented as

Y : 0 5 81 90 70 65 ...

Remark : In fact variable,data etc.can be defined in different way.In this text we keep this definition thinking of the relevant concerned readers.

2.1.2 Frequency Distribution.

Frequency. The number of occurence of an observation or data of a variable is called the frequency of that data.

Illustration. Let x be the marks obtained by 30 students.Let x assumes the values

30	35	31	32	34	31
30	34	42	30	57	68
42	71	20	15	10	51
57	51	51	52	51	80
51	57	20	71	35	32

Here we see the data 30 occurs three times.So frequency of 30 is 3. Similarly the frequency of the data 80,71,68 are 1,2,1 respectively.

Simple Frequency Distribution. The simple frequency distribution of a variable is the statistical table where the observations (assumed by the variable) are arranged in order of magnitude and the frequency of each observation is shown side by side.

Illustration. Let x be a variable which takes the value :

3	4	5	3	6	4
4	3	2	5	6	1
3	4	5	3	2	1

Then the frequency distribution of x is

x :	1	2	3	4	5	6	Total
f_i :	2	2	5	4	3	2	18

The table can also be shown in column-wise.

Grouped Frequency Distribution.

When a large number of datas are available we cannot grasp their characteristic only by placing them individually in a table . In these cases we group the observations into a number of suitable intervals.In a table (or statistical table) these intervals are shown and the frequency of the observations included in each interval are shown side by side.This table is called Grouped frequency distribution.

Illustration. The datas below give the marks secured by 70 candidates in a certain examination :

21	31	35	52	64	74	89	53	42	7
22	35	43	67	76	35	46	26	32	40
72	43	38	41	63	71	28	32	45	54
15	18	52	73	86	50	39	55	47	12
44	58	67	85	39	40	50	65	72	69
57	63	5	56	79	37	24	54	82	49
51	54	68	29	34	44	58	62	59	65

Here we see there is a large number of the observations which are almost distinct.We group the datas into the intervals 0-10,11-21,22-32,....We see the data 7,5 are included in the interval 0-10.So the frequency (called *class-frequency*)of the class interval 0-10 is 2. In this way we have the following grouped frequency distribution :

Marks secured:	0-10	11-21	22-32	33-43	44-54	55-65	66-76	77-87	88-98
Frequency :	2	4	8	14	15	12	10	4	1

Note. Frequency distribution is nothing but quantitative classification.

Terms associated with Grouped Frequency Distribution.

(1) Class interval : The group of datas into a number of suitable intervals are called class interval. In the previous example $0-10$, $11-21$ etc. are class intervals.

(2) Class limits : The two extreme values specifying a class interval are called class limit. In the previous example the lower class limit (LCL) and the upper class limit (UCL) of the class $22-32$ are respectively 22 and 32.

(3) Class Boundaries : The class boundaries of a class are defined as

Lower Class Boundary $(LCB) = LCL$ of the class $-\frac{d}{2}$

where $d = LCL$ of the class $-UCL$ of the previous class.

Upper class Boundary $(UCB) = UCL$ of the class $+\frac{d}{2}$

where $d = LCL$ of the next class $-UCL$ of the class. In the previous example LCB of the class $22-32$ is

$22-\frac{22-21}{2}=21\cdot5$; UCB of the class $22-32$ is

$32-\frac{33-32}{2}=32\cdot5$.

(4) Class Mark or Mid Value : Class Mark of a class

$=\frac{1}{2}(LCL+UCL)$ of the class.

In the previous example, class Mark of the class $66-76$ is $\frac{1}{2}(66+76)=71$.

(5) Width of a class : Width of a class = $(UCB-LCB)$ of the class. In the previous example, width of the class $22-32$ is $32\cdot5-21\cdot5=11$.

Cummulative Frequency.

For a simple frequency distribution the total frequency of the observations lesser or equal to an observation is called "less (≤) than type " cummulative frequency of the observation.

For a grouped frequency distribution the total frequency of observations lesser or equal to the observation in a class is called the "less (≤) than type" cummulative frequency of the observation.

Illustration.

(i) In the simple frequency distribution

x	:	2	4	9	11
f	:	3	6	4	1

the "less than (≤) type" cummulative frequency of 9 is $3+6+4=13$; the "greater than (≥) type" cummulative frequency of 9 is $4+1=5$.

(ii) In the grouped frequency distribution

Class	:	$0-4$	$5-9$	$10-14$	15-19
Frequency	:	3	6	4	1

the "less than (≤) type" cummulative frequency of the class $10-14$ or against the upper boundary $14\cdot5$ of this class is $3+6+4=13$; the "greater than type(≥)" cummulative frequency of the class $10-14$ or against the LCB $9\cdot5$ of this class is $4+1=5$.

2.1.3. Mean.

A typical value which may or may not be among the datas assumed by a variable is considered as a representative of all the datas. For example among the datas 2,5,6.1,8,9,7 the value 7 can be treated as that representative.Generally this representative-value tends to lie centrally within the set of datas .This value is measured by different way. Following is the one of best such measurements.

Arithmetic Mean. The Arithmetic mean (A.M) or briefly the mean of the values (datas) $x_1, x_2, \ldots\ldots x_n$ is defined as

$$\bar{x}=\frac{1}{n}(x_1+x_2+\ldots+x_n)=\frac{1}{n}\sum_{i=1}^{n}x_i.$$

If the datas have the frequency shown in the following table

Variable (x):	x_1	x_2	x_3	...	x_n
Frequency (f_i):	f_1	f_2	f_3	...	f_n

then their A.M is $\bar{x}=\frac{1}{N}(x_1f_1+x_2f_2+\ldots+x_nf_n)=\frac{1}{N}\sum_{i=1}^{n}x_if_i$

where $N=f_1+f_2+\ldots.f_n=\sum_{i=1}^{n}f_i$

For a grouped-frequency distribution the A.M, $\bar{x}=\frac{1}{N}\sum_{i=1}^{n}x_if_i$ where x_i is the mid-value of each class-interval.

Illustration. If

x :	2	4	1	3	Total
f_i :	3	2	4	1	10

be the frequency distribution of a variable x then its A.M,

$$\bar{x}=\frac{1}{10}(2\times3+4\times2+1\times4+3\times1)$$

or , $\bar{x}=\frac{1}{10}(6+8+4+3)=\frac{1}{10}\times21=2.1$

Theorem 1. If the two variables x and y are related by the equation $y=\frac{x-c}{d}$ then $\bar{y}=\frac{\bar{x}-c}{d}$, where c and d are any number.

Proof. We consider the frequency distribution of x :

x:	x_1	x_2	x_3 ...	.... x_n	Total
f_i:	f_1	f_2	f_3 ...	.... f_n	N

Since the values of x are changed to those of y so the frequency distrbution of y would be

y:	y_1	y_2	y_3 ...... y_n	Total
f_i:	f_1	f_2	f_3 ...... f_n	N

where $y_i=\frac{x_i-c}{d}$.

Now the A.M of y, $\bar{y}=\frac{1}{N}\sum_{i=1}^{n}f_i\,y_i=\frac{1}{N}\sum_{i=1}^{n}f_i\frac{x_i-c}{d}$

$$=\frac{1}{Nd}\sum_{i=1}^{n}(f_ix_i-f_ic)=\frac{1}{Nd}\left(\sum_{i=1}^{n}f_ix_i-\sum_{i=1}^{n}f_ic\right)$$

$$=\frac{1}{Nd}\left(\sum_{i=1}^{n}f_ix_i-c\sum_{i=1}^{n}f_i\right)=\frac{1}{Nd}\left(\sum_{i=1}^{n}f_ix_i-cN\right)$$

$$=\frac{1}{Nd}\sum_{i=1}^{n}f_ix_i-\frac{c}{d}=\frac{1}{d}\frac{1}{N}\sum_{i=1}^{n}f_ix_i-\frac{c}{d}=\frac{1}{d}\bar{x}-\frac{c}{d}=\frac{\bar{x}-c}{d}.$$

Note. The above theorem is very much helpful to determine the A.M of the variable assuming large data.This is shown in the following Illustration.

Illustration. We are given the following Grouped frequency distribution :

Weight in gms (x):	110-119	120-129	130-139	140-149	150-159
Frequency :	5	7	12	20	16

Weight in gms (x)	:	160-169	170-179	180-189
Frequency	:	10	7	3

To find the A.M we construct the following table :

Calculation of Mean

Class-interval	Midpoint (x_i)	Frequency (f_i)	x_i -154.5	$\frac{x_i-154.5}{10}=y_i$	f_iy_i
110-119	114.5	5	−40	−4	−20
120-129	124.5	7	−30	−3	−21
130-139	134.5	12	−20	−2	−24
140-149	144.5	20	−10	−1	−20
150-159	154.5	16	0	0	0
160-169	164.5	10	10	1	10
170-179	174.5	7	20	2	14
180-189	184.5	3	30	3	9
Total	-	80	-	-	−52

Here $N=80$ and $\sum_{i=1}^{8}f_iy_i=-52$ .So the

A.M of y, $\bar{y}=\frac{1}{80}\times-52=-0.65$

Since $y_i=\frac{x_i-154.5}{10}$, so $\bar{y}=\frac{\bar{x}-154.5}{10}$

or, $-0.65=\frac{\bar{x}-154.5}{10}$ or, $\bar{x}=148$.

Thus the required A.M, $\bar{x}=148$.

Theorem. 2. (On Mean of Composite Group)

Let $\bar{x}_1,\bar{x}_2,\cdots\cdots\bar{x}_r$ be the *A.M.* of r groups containig $n_1,n_2,\cdots\cdots n_r$ observations respectively. Then their combined mean or composite mean is given by

$$\bar{x}=\frac{n_1\bar{x}_1+n_2\bar{x}_2+\cdots\cdots+n_r\bar{x}_r}{n_1+n_2+\cdots+n_r}$$

Proof. Omitted.

Illustration. Suppose the mean wage of 60 labourers in morning shift is Rs 80 and the mean wage of 40 labourers working in evening shift is Rs 70. Then the mean wage of all labourers (of both shift)

$$=\frac{n_1\bar{x}_1+n_2\bar{x}_2}{n_1+n_2}=\frac{60\times 80+40\times 70}{60+40}=76$$

Note. There are other type of means viz., Geometric mean and Harmonic mean. But study of these is beyond the scope of the book.

2.1.4. Median.

Median of a set of observations is the middle most value when the observations are arranged in increasing or decreasing order of magnitude.

Thus to find the median of a set of observation it is necessrary to arrange the observations in order of magnitude.

Calculation of Median.

Calculation of median may be confusing for even / odd number of observation; for grouped frequency distribution. So we classify the procedure of Calculation of Median in the following three cases :

Case 1. (For simple distribution i.e. without having any frequency).

Arrange the given n number of observations in ascending / descending order of magnitude.

(i) If n is odd then

Median $=\frac{n+1}{2}$ th observation of the arranged set.

(ii) If n is even then

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

Illustration. (i) If we are required to find the median of the set {5, 8, 7, 20, 13, 3, 11} then we first arrange the datas in increasing order i.e. 3, 5, 7, 8, 11, 13, 20. Here the number of observations, $n=7$ (odd).

So its median $=\frac{7+1}{2}$ th $=4$ th observation $=8$.

(ii) If we are required to find the median of the set {5, 4, 7, 3, 21, 12}, we first arrange the datas in ascending order of magnitude {3, 4, 5, 7, 12, 21}. Here the number of observation, $n=6$ (even).

So its median $=\frac{1}{2}\{\frac{6}{2}$ th observation $+\left(\frac{6}{2}+1\right)$ th observation}

$=\frac{1}{2}$ {3rd observation $+4$ th observation} $=\frac{1}{2}(5+7)=6$.

Case 2. (For Simple Frequency Distribution).

Arrange the observations in ascending order of magnitude. Construct "less (≤) than type" cumulative frequency. Calculate $\frac{N+1}{2}$ where N is total frequency. Then Median = the observation corresponding to the cummulative frequency $\frac{N+1}{2}$ or next heigher (if $\frac{N+1}{2}$ is not a cummulative frequency).

Illustration.

Consider the following frequency distribution

Marks :	30	40	50	60	70	80
No of students :	8	15	23	16	8	5

The "less (≤) than type" Cummulative Frequency.

Marks (x) (1)	Frequency (f_i) (2)	Cummulative Frequency (less(≤) than type) (3)
30	8	8
40	15	23
50	23	46
60	16	62
70	8	70
80	5	75
Total	N=75	—

Here $\frac{N+1}{2}=\frac{75+1}{2}=38$.

We see there is no cummulative frequency '38' in column

3. Next higher figure than 38 in column 3 is 46.

$\therefore$ Required Median is the observation corresponding to the cummulative frequency 46 = 50.

Note : $\frac{N+1}{2}$ may be fraction. The procedure is same in that case also.

Case 3. (For Grouped Frequency Distribution)

Here also construct the "less (≤) than type" cummulative frequency against class boundaries. Calculate $\frac{N}{2}$ where N is total frequency. Find the Median-Class, i.e. the class corresponding to the cummulative frequency $\frac{N}{2}$ or next higher (if $\frac{N}{2}$ is not a cummulative frequency).

$$\text{Then Median } = l_m + \frac{\frac{N}{2}-F}{f_m}\times i$$

where l_m = lower boundary of Median-Class

N = Total frequency

F = Cummulative frequency of the class preceeding to the Median-Class.

f_m = frequency of median class.

i = width of the median class.

Illustration.

Consider the following grouped frequency distribution :

Class-interval	Frequency
130–134	5
135–139	15
140–144	28
145–149	24
150–154	17
155–159	10
160–164	1

The "less than type" Cummulative frequency against Class boundaries.

Class-interval (1)	Frequency (2)	Upper Class-Boundary (3)	Cummulative Frequency (≤ type) (4)
130–134	5	134.5	5
135–139	15	139.5	20*
140–144	28	144.5	48
145–149	24	149.5	72
150–154	17	154.5	89
155–159	10	159.5	99
160–164	1	164.5	100
Total	N=100	——	——

* It means there are 20 observations which are less or equal to 139.5.

Now, $\frac{N}{2}=50$. There is no cummulative frequency 50 in the 4th column of the above table. In the 4th column the next higher figure is 72. This corresponds to the median class 145–149.

Therefore, l_m = lower boundary of the median class $=144\cdot5$.

$F =$ the cummulative frequency of the class preceeding the median class $= 48$.

$f_m =$ frequency of the median class $= 24$.

$i =$ width of the median class $= 5$.

$$\text{So the Median } = l_m + \frac{\frac{N}{2} - F}{f_m} \times i = 144.5 + \frac{50-48}{24} \times 5 = 144.92.$$

Note. There is another formula to find the median called "interpolation formula".

2.1.5. Mode.

The observation having maximum frequency is called mode.

Calculation of Mode.

Case 1. For a simple frequency distribution the mode is calculated by simply method of inspection.

Illustration. For the frequency distribution

x :	10	20	30	40	50	70
Frequency :	2	2	3	2	2	1

Here we see the observation 30 has heighest frequency 3. So the mode is 30.

Case 2. (*For Grouped Frequency Distribution*).

For a grouped frequency distribution mode can be calculated for frequency distribution having unique class with heighest frequency and with equal class width. First find the modal class i.e. the class having heighest frequency. Then

$$\text{Mode } = l_m + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

where $l_m =$ lower class boundary of the modal class

$f_1 =$ frequency of the modal class

$f_0 =$ frequency of the class preceeding the modal class

$f_2 =$ frequency of the class succeeding the modal class

$i =$ width of each class (note that this is same for each

Illustration.

Consider the grouped frequency distribution.

Marks	No. of Candidates
0–9	4
10–19	9
20–29	12
30–39	18
40–49	20
50–59	12
60–69	10
70–79	9
80–89	4
90–99	2

Here we see every class has same width which is 10 and only class 40–49 has heighest frequency. So the modal class is 40–49. As we know

$$\text{Mode } = l_m + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

where $l_m =$ lower class boundary of the class 40–49 $= 39.5$

$f_1 =$ frequency of the class "40–49" $= 20$.

$f_0 =$ frequency of the class preceeding the class 40–49 $= 18$

$f_2 =$ frequency of the class succeeding 40–49 $= 12$

$i =$ width of each class $= 10$.

$$\therefore \text{Mode } = 39.5 + \frac{20-18}{40-18-12} \times 10 = 41.5.$$

Relation among Mean, Median and Mode : For a distribution having single Mode the relation is

Mean – Mode ≅ 3(Mean – Median)

Note. For a symetrical distribution mean, median and mode coincide.

2.1.6. Variance and Standard Deviation.

As we have stated in the previous article A.M represents the entire set of datas. But the degree to which the datas tend to spread about the A.M is to be measured.It is usually measured by variance or standard deviation which are discussed below :

Variance.

The mean of the squares of the differences of the observations (or datas) assumed by a variable from their arithmetic mean (A.M) is called variance of the variable.

Standard Deviation.

The positive square root of variance is called Standard Deviation (s.d)

Thus **(i)** if $x_1, x_2, ..., x_n$ be the datas then their variance,

$$Var(x)=\frac{1}{n}\left\{(x_1-\bar{x})^2+(x_2-\bar{x})^2+.....+(x_n-\bar{x})^2\right\}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$ and

the standard deviation, $\sigma_x=+\sqrt{\frac{1}{n}\sum(x_i-\bar{x})^2}$

(ii) if the datas have the frequency shown in the following table

Variable	(x) : x_1	x_2	...	... x_n
Frequency	(f_i) : f_1	f_2	...	... f_n

then the Variance,

$$Var(x)=\frac{1}{N}\left\{f_1(x_1-\bar{x})^2+f_2(x_2-\bar{x})^2+.....+f_n(x_n-\bar{x})^2\right\}$$

$=\frac{1}{N}\sum_{i=1}^{n}f_i(x_i-\bar{x})^2$ where $N=f_1+f_2+....+f_n$ and the standard deviation, $\sigma_x=+\sqrt{\frac{1}{N}\sum_{i=1}^{n}f_i(x_i-\bar{x})^2}$.

If the datas have grouped frequency distribution then x_i will be the mid -value of each class-interval and f_i would be the corresponding frequency.

Illustration : Let x be a variable which assumes the datas: 20,85,120,60,40.Then $\bar{x}=\frac{1}{5}(20+85+120+60+40)=65$. To find the variance and standard deviation we go through the following table:

x_i	x_i-65	$(x_i-65)^2$
20	−45	2025
85	20	400
120	55	3025
60	−5	25
40	−25	625
Total	–	6100

Here $\sum_{i=1}^{5}(x_i-65)^2=6100$

Then $Var(x)=\frac{1}{5}\sum_{i=1}^{5}(x_i-65)^2=\frac{1}{5}\times 6100=1220$.

and the S.d, $\sigma=\sqrt{1220}=34.93$.

Theorem 2. If $x_1, x_2, ... x_n$ be the datas then

$$Var(x)=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2.$$

Proof. Left as an exercise.

Note. The above theorem can be extended for frequency distribution also.There $Var(x)=\frac{1}{N}\sum_{i=1}^{n}f_i x_i^2-\left(\frac{1}{N}\sum_{i=1}^{n}f_i x_i\right)^2$ where $N=f_1+f_2+...+f_n$.

Theorem 3. If the two variables x and y are related by the equation $y=\frac{x-c}{d}$ then $\sigma_y=\frac{\sigma_x}{d}$ where d is a positive number.

Proof. Beyond the scope of this text.

Note. The above theorem is very much helpful to determine the s.d of the variable assuming large data.This is shown in the following illustration.In practice, we use the result in Th-2 to find the variance / S.d.

Illustration. (i) Let we be given the following observation having the frequency distribution :

x :	240.12	240.13	240.15	240.16	240.17	240.21	240.22
Frequency:	2	2	1	1	2	1	1

Scanned by CamScanner

To find the S.d we go through the following table :

x_i	f_i	$x_i-240.16$	$\frac{x_i-240.16}{.01}=y_i$	f_iy_i	$f_iy_i^2$
240.12	2	−.04	−4	−8	32
240.13	2	−.03	−3	−6	18
240.15	1	−.01	−1	−1	1
240.16	1	0	0	0	0
240.17	2	.01	1	2	2
240.21	1	.05	5	5	25
240.22	1	.06	6	6	36
Total	10	-	-	−2	114

Here, $N=10$, $\sum_{i=1}^{7} f_iy_i=-2$ and $\sum_{i=1}^{7} f_iy_i^2=114$

Now, $Var(y)=\frac{\sum f_iy_i^2}{N}-\left(\frac{\sum f_iy_i}{N}\right)^2=\frac{114}{10}-\left(\frac{-2}{10}\right)^2=11.36$

$\therefore \sigma_y=\sqrt{11.36}=3.37$

Since $y_i=\frac{x_i-240.16}{.01}$, so $\sigma_y=\frac{\sigma_x}{.01}$

or, $3.37=\frac{\sigma_x}{.01}$ $\quad\therefore \sigma_x=0.0337$

(ii) Consider the following grouped frequency distribution:

Value :	90–99	80–89	70–79	60–69	50–59	40–49	30–39
Frequency:	2	12	22	20	14	4	1

To find the variance and standard deviation of this grouped frequency distribution we construct the following table :

Class interval	Mid point (x)	frequency (f)	$x-64\cdot5$	$y=\frac{x-64\cdot5}{10}$	fy	fy^2
90–99	94·5	2	30	3	6	18
80–89	84·5	12	20	2	24	48
70–79	74·5	22	10	1	22	22
60–69	64·5	20	0	0	0	0
50–59	54·5	14	−10	−1	−14	14
40–49	44·5	4	−20	−2	−8	16
30–39	34·5	1	−30	−3	−3	9
Total	—	75	—	—	27	127

Here $N=75$, $\sum f_iy_i=27$, $\sum f_iy_i^2=127$.

Now, $\text{Var}(y)=\frac{1}{N}\sum f_iy_i^2-\left(\frac{1}{N}\sum f_iy_i\right)^2=\frac{1}{75}\times127-\left(\frac{27}{75}\right)^2=1{\cdot}56.$

$\therefore$ the standard deviation of y, $\sigma_y=\sqrt{1\cdot56}=1\cdot23$.

Now since $y=\frac{x-64\cdot5}{10}$; therefore $\sigma_y=\frac{\sigma_x}{10}$

or, $\sigma_x=1\cdot23\times10=12\cdot3$.

Illustrative Examples.

Ex. 1. *Find the mean from the following data :*

Daily wages (Rs) :	*25–29*	*30–34*	*35–39*	*40–44*
No. of workers :	*16*	*28*	*14*	*12*

It is given that the total wage for 10 workers earning Rs 45 and more is Rs 600.

First we are to work out the mean for the rest part without the last class. For that we construct the following table :

class interval	mid point (x_i)	frequency f_i	$y_i=\frac{x_i-37}{5}$	f_iy_i
25–29	27	16	−2	−32
30–34	32	28	−1	−28
35–39	37	14	0	0
40–44	42	12	1	12
Total		70		−48

$\therefore$ Here $N = 70, \sum f_i y_i = -48$

$$\therefore \quad \bar{y} = \frac{-48}{70} = \frac{-24}{35}$$

But $\bar{y} = \bar{x} - \frac{37}{5}$

i.e., $\bar{x} = 37 + 5\bar{y} = 37 - \frac{24}{35} \times 5$

$$= 33 \cdot 57 \cdot$$

So the total wage of 70 workers = Rs. $33 \cdot 57 \times 70$

= Rs. $2349 \cdot 9$ ·

Thus the wage of total 80 workers is

Rs $(2349 \cdot 9 + 600)$

= Rs $2949 \cdot 9$ ·

$\therefore$ Mean wage is Rs $\frac{2949 \cdot 9}{80}$

= Rs $36 \cdot 87$ ·

Ex. 2. *The A.M calculated from the following frequency distribution is known to be 72.5. Find the value of x :*

classes	*: 30–39*	*40–49*	*50–59*	*60–69*	*70–79*	*80–89*	*90–99*
Frequency :	*2*	*3*	*11*	*20*	*x*	*25*	*7*

We first construct the following table to find A.M using the given datas :

Classes	Mid pt (x_i)	Frequency (f_i)	$y_i = \frac{x_i - 74 \cdot 5}{10}$	$y_i f_i$
30–39	34.5	2	–4	–8
40–49	44.5	3	–3	–9
50–59	54.5	11	–2	–22
60–69	64.5	20	–1	–20
70–79	74.5	x	0	0
80–89	84.5	25	1	25
90–99	94.5	7	2	14
Total		67+x		–20

$$\therefore \quad \bar{y} = \frac{\sum y_i f_i}{\sum f_i} = \frac{-20}{67 + x}$$

$$\therefore \quad \bar{x} = 74.5 + 10\bar{y} = 74.5 - \frac{200}{67 + x} \cdot$$

By the give condition we have.

$$74.5 - \frac{200}{67 + x} = 72.5$$

or, $67 + x = 100$

$\therefore \quad x = 33$ ·

Thus the missing frequency x is 33.

Ex. 3. *Average marks obtained by a class of 70 students was found to be 65. Later it was found that the marks of one student was wrongly recorded as 85 in place of 58. Find the corrected mean.*

Wrongly calculated mean $= \frac{1}{70}\sum_{i=1}^{70} n_i = 65$

or, $\sum_{i=1}^{70} n_i = 4550$

i.e. sum of wrong observations $= 4550$ ·

$\therefore$ sum of corrected observations $= 4550 - 85 + 58 = 4523$ ·

So the corrected mean $= \frac{4523}{70} = 64 \cdot 61$ ·

Ex. 4. *Following is a frequency distribution lacking two class frequency. Find them if the mean is* $7 \cdot 74$ ·

value	*: 3–5*	*5–7*	*7–9*	*9–11*	*11–13*	*Total*
frequency :	*32*	—	*57*	—	*25*	*200*

Let the two missing frequencies be f_1 and f_2 respectively.

We construct the following table :

Class-interval	Frequency	Class-Mark	xf
3–5	32	4	128
5–7	f_1	6	$6f_1$
7–9	57	8	456
9–11	f_2	10	$10f_2$
11–13	25	12	300
Total	200=N	——	$884 + 6f_1 + 10f_2$

Now, $32+f_1+57+f_2+25=200$

or, $f_1+f_2=86$ ... (1)

The mean, $\bar{x}=\frac{1}{N}\sum x_i f_i$

or, $7 \cdot 74=\frac{1}{200}\times(884+6f_1+10f_2)$

or, $3f_1+5f_2=332$ ... (2)

Solving (1) and (2) we get $f_1=49$, $f_2=37$ so the two missing frequencies are 49 and 37.

Ex. 5. *Two variables x and y are related by* $3x+4y=21$. *A. M. of x is 3.*

Find A. M. of y.

From the given relation we get $y=\frac{21-3x}{4}$.

$\therefore\ \bar{y}=\frac{21-3\bar{x}}{4}=\frac{21-3\times 3}{4}=\frac{12}{4}=3$.

Ex. 6. *Two variables x and y are related by* $x=2y+5$. *The median of x is 25. Find the median of y.*

From the given relation we have $y=\frac{x-5}{2}$.

Here if x increases y also does so.

So the median of $y=\frac{\text{Median of } x-5}{2}=\frac{25-5}{2}=10$.

Ex. 7. *The number of observations of two groups are in the ratio 2:1 and their A. M. are 8 and 128 respectively. Find the A. M. of the combined group.*

Let the number of observations of the two groups be $2k,k$. $\bar{x}_1=8$, $\bar{x}_2=128$. The combined A. M.,

$$\bar{x}=\frac{n_1\bar{x}_1+n_2\bar{x}_2}{n_1+n_2}=\frac{2k\times 8+k\times 128}{2k+k}$$

$$\text{or, } \bar{x}=\frac{144k}{3k}=48.$$

Ex. 8. *Find the median of the following frequency distribution :*

x :	5	10	15	20	25	30	35	40
f :	7	10	15	18	23	21	17	8

The cumulative frequency distribution table is given below :

x	f	c. f
5	7	7
10	10	17
15	15	32
20	18	50
25	23	73
30	21	94
35	17	111
40	8	119

Here $N=119$

$\therefore\ \frac{N+1}{2}=60$.

So the cumulative frequency just greater than $\frac{N+1}{2}$ is 73 and the value of x corresponding to c.f 73 is 25. Hence the median is 25.

Ex. 9. *Calculate mean, median and hence find the approximate value of the mode from the following frequency distributions :*

Height (inches):	*60–63*	*64–67*	*68–71*	*72–75*	*76–79*	*80–83*
No of students :	*8*	*3*	*18*	*6*	*16*	*8*

To find mean we first construct the following table :

classes	Mid pt (x_i)	frequency (f_i)	c.f	$y_i = \frac{x_i - 73\cdot5}{4}$	$y_i f_i$
60–63	61.5	8	8	–3	–24
64–67	65.5	3	11	–2	–6
68–71	69.5	18	29	–1	–18
72–75	73.5	6	35	0	0
76–79	77.5	16	51	1	16
80–83	81.5	8	59	2	16
					–16

Here $N = \sum f_i = 59$

$$\therefore \quad \bar{y} = \frac{\sum y_i f_i}{\sum f_i} = \frac{-16}{59} = -0.2712$$

Now, $y_i = \frac{x_i - 73.5}{4}$ or, $x_i = 73.5 + 4y_i$

$$\therefore \quad \bar{x} = 73.5 + 4\bar{y} = 73.5 + 4(-0.2712)$$
$$= 72.4152$$

$\therefore$ mean of the given distribution $= 72.4152$.

Now $\frac{N}{2} = 29.5$.

So the median class is 72–75

$\therefore \; l_m = 71.5, F = 29, f_m = 6$

and $i = 75.5 - 71.5 = 4$

$$\therefore \text{ Median } = lm + \frac{\frac{N}{2} - F}{f_m} \times i$$

$$= 71.5 + \frac{29.5 - 29}{6} \times 4 = 71.83 .$$

Using the relation Mean-Mode $= 3$ (Mean-Median) we have

72.4152–Mode $= 3(72.4152 - 71.83)$ or, 72.4152–mode $= 1.7556$

$\therefore$ Mode $= 70.6596$

Ex. 10. *Calculate the mode of the following data :*

1, 12, 5, 8, 12, 13, 8, 1, 4, 8, 7, 8, 5.

Let us arrange the given variaties with corresponding frequencies as given below

x	f
1	2
4	1
5	2
7	1
8	4
12	2
13	1

As the variate 8 ocurs 4 times which is maximum, so, mode $= 8$.

Ex. 11. *Calculate the mode from the following distribution :*

class	:	*10–15*	*15–20*	*20–25*	*25–30*	*30–35*
frequency	:	*6*	*9*	*11*	*7*	*7*

Here the greatest frequency 11 lies in the class 20–25. Hence modal class is 20–25.

$\therefore \; l_m =$ lower class boundary of the modal class $= 20$.

$f_1 =$ frequency of the class $= 11$

$f_0 =$ frequency of the procceding the modal class $= 9$.

$f_2 =$ frequency of the class succeding the modal class $= 7$.

$i =$ width of each class $= 6$.

$$\therefore \text{ Mode } = l_m + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 20 + \frac{11 - 9}{22 - 9 - 7} \times 6 \quad = 22 .$$

Ex.12. *The median and mode of the following frequency distribution are known to be 27 and 26 respectively. Find the values of α and β :*

Scanned by CamScanner

class interval: 0–10 10–20 20–30 30–40 40–50
frequency : 3 α 20 12 β

Since mode $=26$, so it lies within the class 20–30.

$$\text{Mode} = l_m + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

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

$$= 20 + \frac{20-\alpha}{28-\alpha} \times 10.$$

$$\therefore \quad 20 + \frac{10(20-\alpha)}{28-\alpha} = 26$$

or, $\frac{10(20-\alpha)}{28-\alpha} = 6$

$\therefore \quad \alpha = 8$.

So the total frequency is $N = 43 + \beta$.

But media $=27$, so it lies within the class 20–30.

Then we construct the following frequency table :

class interval	frequency	*c.f.* frequency
0–10	3	3
10–20	8	11
20–30	20	31
30–40	12	43
40–50	β	$43+\beta$

Now median $= l_m + \dfrac{\frac{N}{2} - F}{f_m} \times i$

$$\therefore \quad 27 = 20 + \frac{\frac{43+\beta}{2} - 11}{20} \times 10$$

or, $\frac{21+\beta}{4} = 7$

or, $21 + \beta = 28$

$\therefore \quad \beta = 7$.

Ex. 13. *Find the variance and standard deviation of the following frequency distribution :*

Weight (in kg) : 36–40 41–45 46–50 51–55 56–60 61–65 66–70

No of persons : 14 26 40 33 50 37 25

We construct the following table :

class inteval	Mid pt (x_i)	frequency f_i	$y_i = \frac{x_i - 53}{5}$	$f_i y_i$	$f_i y_i^2$
36–40	38	14	–3	–42	126
41–45	43	26	–2	–52	104
46–50	48	40	–1	–40	40
51–55	53	33	0	0	0
56–60	58	50	1	50	50
61–65	63	37	2	74	148
66–70	68	25	3	75	225
Total		225		65	693

Here $N = 22, \sum f_i y_i = 65, \sum f_i y_i^2 = 693$

$$\therefore \quad Var(y) = \frac{1}{N} \sum f_i y_i^2 - \left(\frac{1}{N} \sum f_i y_i\right)^2$$

$$= \frac{1}{225} \times 693 - \left(\frac{65}{225}\right)^2$$

$$= 2.996.$$

Scanned by CamScanner

$\therefore$ S.D of y, $\sigma_y = \sqrt{2.996} = 1.731$

Since $y_i = \dfrac{x_i - 53}{5}$, so $\sigma_y = \dfrac{\sigma_x}{5}$

$\therefore \quad \sigma_y = 1.731 \times 5 = 8.655$.

Ex. 14. *The mean and standard deviation of marks of 70 students were found to be 65 and 5.2 respectively. Later it was detected that the marks of one student was wrongly recorded as 85 instead of 58. Obtain the correct s.d.*

Let $x_1, x_2, x_3 \cdots$ be the marks.

The "incorrect $\sum x_i$ " $= 65 \times 70 = 4550$.

$\therefore$ the "correct $\sum x_i$ " $= 4550 - 85 + 58 = 4523$.

$\therefore$ the 'correct mean' $= \dfrac{4523}{70} = 64 \cdot 61$.

We know $(s.d)^2 = \dfrac{1}{n}\sum x_i^2 - (\bar{x})^2$ or, $\sum x_i^2 = n\left((s.d)^2 + \bar{x}^2\right)$.

Again the "incorrect $\sum x_i^2$ " $= 70\left\{(5 \cdot 2)^2 + (65)^2\right\} = 297642 \cdot 8$.

Then the correct $\sum x_i^2 = 297642 \cdot 8 - 85^2 + 58^2 = 293781 \cdot 8$.

Hence the correct $(s.d)^2 = \dfrac{293781 \cdot 8}{70} - (64 \cdot 61)^2 = 22 \cdot 43$.

So the correct $s.d = \sqrt{22 \cdot 43} = 4 \cdot 74$.

Ex. 15. *Three factories A, B, C producing similar products are such that the mean daily wage of workers of factory A is Rs. 100 with a s.d of Rs. 10, whereas in factory B, the mean wage is Rs. 150 and s.d is Rs. 12 and in factor C, the mean wage is Rs. 150 and s.d is Rs. 10. Which factory is most consistent in respect of the daily wage of their workers ?*

The s.d of workers of factory A per unit-mean $= \dfrac{10}{100} = 0 \cdot 1$.

The s.d of workers of factory B per unit-mean $= \dfrac{12}{150} = \cdot 08$.

The s.d of workers of factory C per unit-mean $= \dfrac{10}{150} = 0 \cdot 07$.

This shows that the daily wage of the workers of factory C is most consistant. Variability is highest for factory A.

EXERCISES

1. Find the A.M of the variable assuming the datas 8, 5, 3, 10, 12.

2. Find the A.M of the following frequency distribution :

x	:	6	5	2	8
Frequency	:	4	3	1	2

3. The scores of a cricketer playing six matches are 84, 91, 72, 68, 87 and 78. Find the arithmetic mean (A.M) of the scores.

4. Ten measurements of the volume of a cone were recorded by an engineer as 3.88, 4.09, 3.92, 3.97, 4.02, 3.95, 4.03, 3.92, 3.98 and 4.06 c.c.Find the A.M of the measurements.

5. Find the A.M of the following frequency distribution :

x	:	462	480	498	516	534	552	570	588	606	624
f	:	98	75	56	42	30	21	15	11	6	2

6. Following is the frequency distribution for the number of minutes per week spent watching TV by 400 senior citizens.

Viewing Time (minute)	:	300-399	400-499	500-599	600-699	700-799
Number of Citizens	:	14	46	58	76	68

	800-899	900-999	1000-1099	1100-1199
	62	48	22	6

Find the mean TV viewing time for the 400 senior citizens per week.

7. Four groups of cattles,consisting of 18, 10, 20, and 15 cattles,reported weight are 140, 153, 148 and 162 Kg respectively.Find the mean weight of all the cattles.
[Hint : It is a frequency distribution like

x	:	140	153	148	162
Frequency	:	18	10	20	15]