# STATISTICS

A collection of methods for collecting, displaying,analyzingand drawing conclusions from data.

MEASURES OF CENTRAL TENDENCY

One of the powerful tools of analysis data is to calculate a single average value that represents the entire mass of data. The word average is very commonly used in day-to-day conversation. An 'average' thus is a single value which is considered as the most representative or typical value for a given set of data. Such a value is neither smallest nor largest value, but is a number whose value is somewhere in the middle of the group. For this reason an average is frequently referred to as a measure of central tendency. Measures of central tendency show the tendency of some central value around which data tends to cluster.

The following are the important measures of central tendency:
A. Arithmetic Mean   B. Median   C. Mode
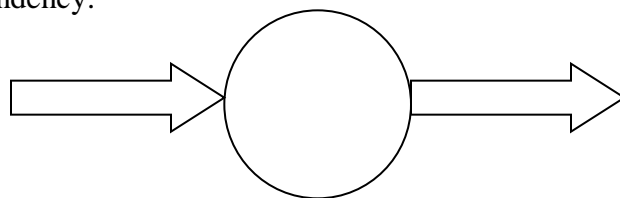
A. Arithmetic Mean

For Ungrouped Data

If $X_1, X_2, X_3, \ldots, X_N$ represent the values of N observations, the arithmetic mean denoted by $\bar{X}$ is

defined as: $\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_N}{N} = \frac{\sum_{i=1}^{N} X_i}{N}$

For Grouped Data (Direct Method)

If $X_1, X_2, X_3, \ldots, X_N$ represent the values of N observations, the arithmetic mean denoted by $\bar{X}$ is

defined as: $\bar{X} = \frac{f_1 X_1 + f_2 X_2 + f_3 X_3 + \cdots + f_N X_N}{N} = \frac{\sum_{i=1}^{N} f_i X_i}{N}$

$X = \text{mid} - \text{point of various classes}$
$f = \text{frequency of each class}$
$N = \text{total frequncy}$

For Grouped Data (Short-cut Method)

$\bar{X} = A + \frac{\sum fd}{N} \times i \qquad d = \frac{X - A}{i}$

$A = \text{Assumed Mean}$
$N = \text{total frequncy}$
$i = \text{size of class interval}$

Observations......Profit, Salary, Age, Height, Marks, Income etc......X
Frequency....... Number of students, companies, workers, people etc...... f

Problem.

The following are the figures of profits earned by 14,00 companies:

Profits (in Lakhs)	No of Companies
200—400	500
400—600	300
600—800	280
800—1000	120
1000—1200	100
1200—1400	80
1400—1600	20

Calculate the **average profit** for all the companies.

Solution.

Direct Method:

$$Average\ profit = \frac{Total\ profit}{Total\ No.\ of\ companies}$$

$$\bar{x} = \frac{\sum fX}{N}$$

Profits (in lakhs)	Midpoint X	No of Companies f	fX	
200—400	300	500	150000	
400—600	500	300	150000	
600—800	700	280	196000	
800—1000	900	120	108000	
1000—1200	1100	100	110000	
1200—1400	1300	80	104000	
1400—1600	1500	20	30000	
		$N = 1400$	$\sum fX$ $= 848000$	

$$\bar{x} = \frac{\sum fX}{N} = \frac{848000}{1400} = 605.71$$

Short-cut Method

$$\bar{X} = A + \frac{\sum fd}{N} \times i \qquad d = \frac{X-A}{i}$$

A = Assumed Mean $\qquad$ N = total frequncy $\qquad$ i = size of class interval

Profits (in lakhs)	Midpoint X	No of Companies f	$d = \dfrac{X-A}{i}$	fd
200—400	300	500	−3	-1500
400—600	500	300	-2	-600
600—800	700	280	-1	-280
800—1000	900=A	120	0	0
1000—1200	1100	100	1	100
1200—1400	1300	80	2	160
1400—1600	1500	20	3	60
		$N = 1400$		$\sum fd = -2060$

$$\bar{X} = A + \frac{\sum fd}{N} \times i = 900 - \frac{2060}{1400} \times 200 = 605.71$$

MEDIAN

The median is the measure of central tendency which appears in the "middle" of an ordered squence of values. Hat is half of the observation in a set of data are lower than it and half of the observations are greater than it.

<div align="center">2,2,2,3,5,5,5,5,5,7,7,7,8,8,8,8,8,10,10</div>

Note: The arithmetic mean is calculated from the value of every observation in the series whereas **median is called a positional average**.

Calculation of Median (For Grouped Data)

Determine the particular case in which the median value lies. Use $\frac{N}{2}$ to locate median class but not $\frac{N+1}{2}$ because in grouped data, $\frac{N}{2}$ divides the area of the curve into two equal parts.

Formula of Median

$$\textbf{Median} = L + \frac{N/2 - pcf}{f} \times i$$

where L = Lower limit of median class

f = Frequency of median class

i = $Size\ of\ c$lass interval of median class

pcf = Preceding Cumulative frequency of median class.

cumulative frequency$(cf) \dots \dots \dots \dots \dots$ When frequencies are added, the new frequencies are known as cf
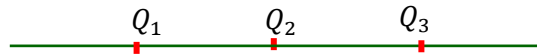
The procedure for computing quartiles, deciles, etc., is the same as for median.

Formulae of quartiles, deciles, percentiles;

$$Q_j = L + \frac{jN/4 - pcf}{f} \times i \qquad \text{for } j = 1, 2, 3$$

$$D_k = L + \frac{kN/10 - pcf}{f} \times i \qquad \text{for } k = 1, 2, \dots \dots \dots \dots, 9$$

$$P_l = L + \frac{lN/100 - pcf}{f} \times i \qquad \text{for } l = 1, 2, \dots \dots \dots \dots, 99$$

$Q_1 \qquad Q_2 \qquad Q_3$

Problem.

The following are the figures of profits earned by 14,00 companies:

Profits (in Lakhs)	No of Companies
200—400	500
400—600	300
600—800	280
800—1000	120
1000—1200	100
1200—1400	80
1400—1600	20

Calculate the **median profit** for all the companies. Also find central 50% profit.

Central 50% $= \boldsymbol{Q_3 - Q_1}$

Solution.

cumulative frequency(cf) … … … … … When frequencies are added, the new frequencies are known as cf

Profits (Tk. in Lakhs)	No. of Companies(f)	Profits (Tk. in Lakhs)	No. of Companies(cf)
200—400	500	Less than 400	500
400—600 Me class	300	Less than 600	800
600—800 Q_3, D_7 class	280	Less than 800	1080
800—1000 P_{80} class	120	Less than 1000	1200
1000—1200 D_9 class	100	Less than 1200	1300
1200—1400	80	Less than 1400	1380
1400—1600	20	Less than 1600	1400
	$N=1400$		

$$N/2 = 700$$

$$\textbf{Median} = L + \frac{N/2 - pcf}{f} \times i = 400 + \frac{700 - 500}{300} \times 200 = 533.33$$

$$3N/4 = 1050$$

$$Q_j = L + \frac{jN/4 - pcf}{f} \times i \qquad \text{for } j = 1,2,3$$

$$Q_3 = L + \frac{3N/4 - pcf}{f} \times i = 600 + \frac{1050 - 800}{280} \times 200 = 778.57$$

$$7N/10 = 980$$
$$9N/10 = 1260$$

$$D_k = L + \frac{kN/10 - pcf}{f} \times i \qquad \text{for } k = 1, 2, … … … … , 9$$

$$D_7 = L + \frac{7N/10 - pcf}{f} \times i = 600 + \frac{980 - 800}{280} \times 200 =$$

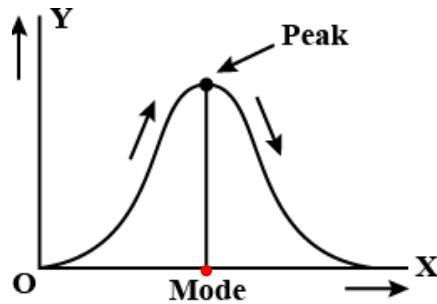$$D_9 = L + \frac{9N/10 - pcf}{f} \times i = 1000 + \frac{1260 - 1200}{100} \times 200 =$$

$$P_l = L + \frac{lN/100 - pcf}{f} \times i \qquad \text{for } l = 1, 2, … … … … , 99$$

$$80N/100 = 1120$$

$$P_{80} = L + \frac{80N/100 - pcf}{f} \times i = 800 + \frac{1120 - 1080}{120} \times 200 =$$

MODE

Mode is defined as that value which occurs the maximum numbers of times i.e. having the maximum frequency. Graphically it is the value on the X−axis below the peak of the frequency curve as can be seen from the following diagram.

Marks	No. of students
4	1
3	3
0	1
2	1
5	2
6 Modal value	4
7	1
8	2
9	1
10	1

Calculation of Mode (For Grouped Data)
Formula of Mode

$$\text{Mo} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i \qquad \textbf{OR} \quad \text{Mo} = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

where
$\Delta_1 =$ Difference between frequency of modal class & $frequency\ of\ pre-modal\ class$
$\Delta_2 =$ Difference between frequency of modal class & $frequency\ of\ post-modal\ class$
$$L = \text{Lower limit of modal class}$$
$i =$ Size of modal class

Note: While applying above formula, it is necessary to see that the class intervals are uniform throughout. If they are unequal, should be equal first on the assumption, otherwise we will get misleading results.

If a distributation contains more than one mode (multimodal distribution), known as ill-defined mode then the mode cannot be determined by nthe above formula. Thus mode is ascertained by the following formula based upon the relationship between mean, median and mode.

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

Problem 1. *The following are the figures*	
of profits earned by 14,00 companies:	
Profits (in Lakhs)	No of Companies
200—400	500
400—600	300
600—800	280
800—1000	120
1000—1200	100
1200—1400	80
1400—1600	20

Problem 2. The weights of 150	
students are classified as follows:	
Weights	No. of students
less than 40	18
less than 50	55
less than 60	100
less than 70	127
less than 80	142
less than 90	150

Calculate the ***modal profit/ weight*** for all the *companies/students*.

Solution.

Profits (Tk. In Lakhs)	No. of Companies f
200—400 Modal Class	500
400—600	300
600—800	280
800—1000	120
1000—1200	100
1200—1400	80
1400—1600	20
	N=1400

$$\text{Mo} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$
$$= 200 + \frac{500 - 0}{(500 - 0) + (500 - 300)} \times 200$$

Weights	No. of students cf	Weights	No. of students f
less than 40	18	30—40	18
less than 50	55	40—50	37
less than 60	100	50—60 Modal Class	45
less than 70	127	60—70	27
less than 80	142	70—80	15
less than 90	150	80—90	8

$$\text{Mo} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$
$$= 50 + \frac{45 - 37}{(45 - 37) + (45 - 27)} \times 10$$

STANDARD DEVIATION

The standard deviation is a measure of how much "spread" or "variability" is present in the sample. If all numbers in the sample are very close to each other, the standard deviation is close to zero. If the numbers are well dispersed, the standard deviation will tend to be large. A small standard deviation means a high degree of uniformity of the observations as well as homogeneity of a series. Hence the standard deviation is extremely useful in judging the representativeness of the mean. Standard deviation is denoted by the small Greek letter sigma (σ) and is defined as:

$$\sigma = \sqrt{\frac{\Sigma(X-\bar{X})^2}{N}}, \bar{X} \text{ is arithmetic mean.}$$

When deviations are taken from assumed mean (A), then

$$\boldsymbol{\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i,} \qquad d = \frac{X-A}{i}$$

If we square standard deviation, we get variance. Hence Variance $= \sigma^2$.

$$\boldsymbol{Average\ Deviation = \frac{\Sigma f|X - \bar{x}|}{N}}$$

$$\boldsymbol{Standard\ Deviation, \sigma = \sqrt{\frac{\Sigma f(X - \bar{x})^2}{N}}}$$

$$\boldsymbol{\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i,}$$

The following are the figures of profits earned by 14,00 companies:

Profits (in Lakhs)	No of Companies
200—400	500
400—600	300
600—800	280
800—1000	120
1000—1200	100
1200—1400	80
1400—1600	20

1. Calculate the **average profit, median profit, modal profit** for all the companies
2. Calculate the **deviation** from average profit.
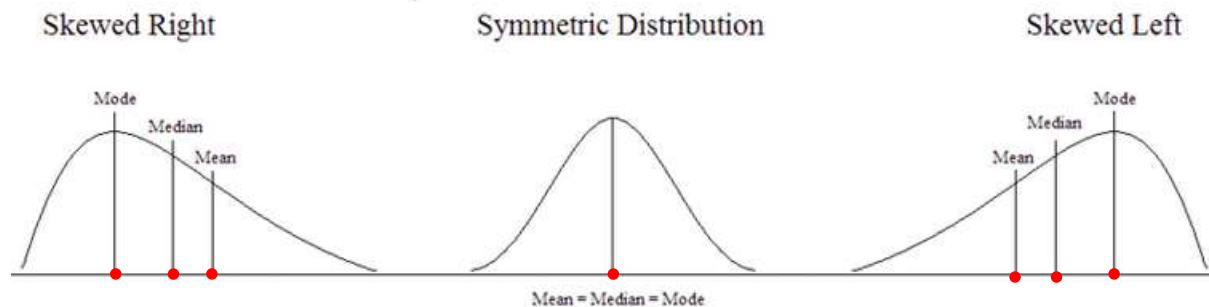
Solution.

Profits (Tk. In Lakhs)	Mid-points X	No. of Companies(f)	d	Profits (Tk. In Lakhs)	No. of Companies(cf)	fd	fd^2
200—400	300	500	-3	Less than 400	500	-1500	4500
400—600 Me Class	500	300	-2	Less than 600	800	-600	1200
600—800 Q_3, D_7 Class	700	280	-1	Less than 800	1080	-280	280
800—1000 P_{80} Class	900=A	120	0	Less than 1000	1200	0	0
1000—1200	1100	100	1	Less than 1200	1300	100	100
1200—1400	1300	80	2	Less than 1400	1380	160	320
1400—1600	1500	20	3	Less than1600	1400	60	180
		N=1400				$\sum fd$ $= -2060$	$\sum fd^2$ $= 6580$

1. Done before

2. $\sigma = \sqrt{\dfrac{\sum fd^2}{N} - \left(\dfrac{\sum fd}{N}\right)^2} \times i = \sqrt{\dfrac{6580}{1400} - \left(\dfrac{-2060}{1400}\right)^2} \times 200 = \sqrt{4.7 - (-1.47)^2} \times 200$

$= \sqrt{4.7 - 2.1609} \times 200 = \sqrt{2.539} \times 200$

$= \sqrt{2.539} \times 200 = 318.68$

SKEWNESS
Lack of symmetry

Skewed Right · Symmetric Distribution · Skewed Left

Mean = Median = Mode

Karl Pearson's coefficients of Skewness
$$SK_p = \frac{Mean - Mode}{\sigma}$$

Bowley's coefficients of Skewness
$$SK_B = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{Q_3 + Q_1 - 2\,\text{Med}}{Q_3 - Q_1}$$

Illustration

The weights of 150 students are classified as follows:

Weight(in kg)	30—40	40—50	50—60	60—70	70—80	80—90
No.of students	18	37	45	27	15	08

1. Find the average and modal weights. Also show the deviation from mean.
2. Calculate Q_1, Median, D_7, P_{80} and interpret the values. Also find central 50% weights.
3. Find Mo, Q_1, Median, D_7, P_{80} from graph and Verify your results.

Solution

Calculation of $\bar{x}, \sigma, \text{Mo}, Q_1, \text{Median}, D_7, P_{80}$

Weight(in Kg)	X	No. of students (f)	$d = \dfrac{X-A}{i}$	fd	fd^2	Weights less than—	No. of students (cf)
30—40	35	18	-2	-36	72	40	18
40—50 Q_1	45	37	-1	-37	37	50	55
50—60 M_O, Me	55=A	45	0	0	0		100
60—70 D_7, P_{80}	65	27	1	27	27		127
70—80	75	15	2	30	60		142
80—90	85	08	3	24	72	90	150
		N=150		$\sum fd$=8	$\sum fd^2$=268		

1. We know $\bar{x} = A + \dfrac{\sum fd}{N} \times i$

$$= 55 + \dfrac{8}{150} \times 10$$

$$= 55.53 \ \text{kg}$$

where A = Assumed Mean
N = total frequncy
i = size of class interval
$d = \dfrac{X-A}{i}$

$M_O = L + \dfrac{\Delta_1}{\Delta_1 + \Delta_2} \times i$

where

$\Delta_1 =$ Difference between frequency of modal class & frequency of pre − modal class
$\Delta_1 =$ Difference between frequency of modal class & frequency of post − modal class $L =$ Lower limit of modal class
$i =$ Size of modal class

Since 45 is the height frequency, so **50—60** is the modal class

$\therefore M_O = 55 + \dfrac{45-37}{(45-37)+(45-27)} \times 10 = 55 + \dfrac{8}{18} \times 10 = 59.44\text{kg}$

$\sigma = \sqrt{\dfrac{\sum fd^2}{N} - \left(\dfrac{\sum fd}{N}\right)^2} \times i = \sqrt{\dfrac{268}{150} - \left(\dfrac{8}{150}\right)^2} \times 10 = \sqrt{1.79 - 0.002} \times 10$

$= \sqrt{1.788} \times 10 = 13.37$

2. Median $= L + \frac{N/2 - pcf}{f} \times i$

where $L = $ Lower limit of median class

$f = $ Frequency of median class

$i = Size\ of\ c$lass interval of median class

$pcf = $ Preceding Cumulative frequency of median class.

Median $= 50 + \frac{75 - 55}{45} \times 10 = 54.44$

$Q_1 = L + \dfrac{N/4 - pcf}{f} \times i$

$D_7 = L + \dfrac{7N/10 - pcf}{f} \times i$
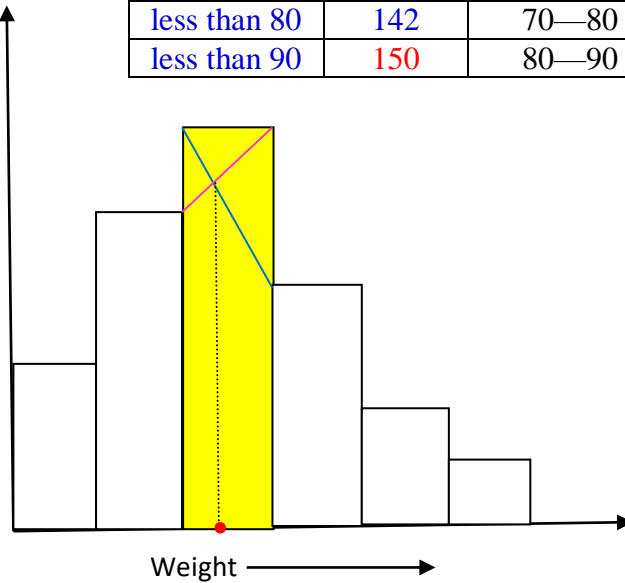
$P_{80} = L + \dfrac{80N/100 - pcf}{f} \times i$

$Q_3 = L + \dfrac{3N/4 - pcf}{f} \times i$
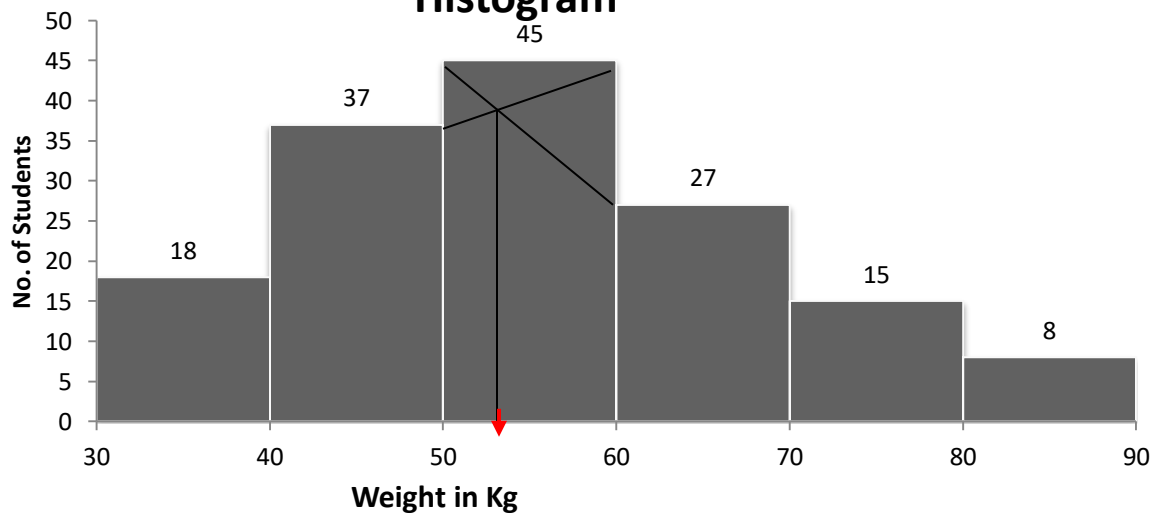
Finding Mode from **Histogram**

Weights	No. of students cf
less than 40	18
less than 50	55
less than 60	100
less than 70	127
less than 80	142
less than 90	150

Weights	No. of students cf	Weights	No. of students f
less than 40	18	30—40	18
less than 50	55	40—50	37
less than 60	100	50—60 Modal Class	45
less than 70	127	60—70	27
less than 80	142	70—80	15
less than 90	150	80—90	8

$$Mo = 50 + \frac{45-37}{(45-37)+(45-27)} \times 10 = 53.07$$

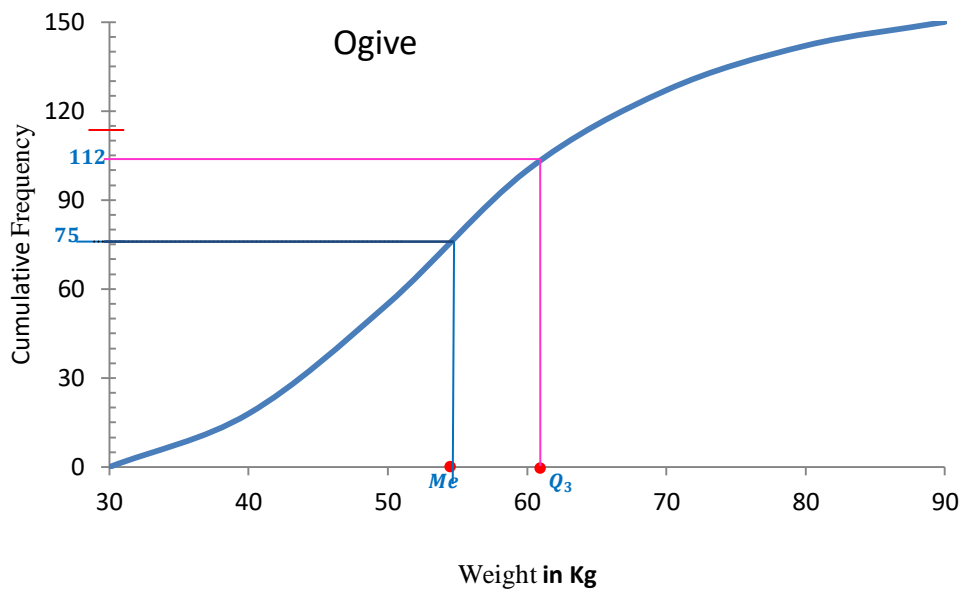Weight →

Histogram

No. of Students vs **Weight in Kg**

In Fig. **Red arrow** indicate modal value.

Finding Median, Quartiles , Desiles, Percentiles from Ogive

Weights	No. of students cf
less than 40	18
less than 50	55
less than 60	100
less than 70	127
less than 80	142
less than 90	150

Weights	No. of students f
30—40	18
40—50	37
50—60	45
60—70	27
70—80	15
80—90	8

Ogive

Red dot indicates Median value.

Problem 1.

The profits of 125 companies are classified as follows:

Profits	No. of companies	Profits	No. of companies
less than 10	4	less than 50	96
less than 20	16	less than 60	112
less than 30	40	less than 70	120
less than 40	76	less than 80	125

1. Find the average and modal profits. Also calculate the deviation from mean.
2. Calculate Q_1, Median, D_7, P_{80} and interpret the values. Also find central 50% weights.
3. Find Mo, Q_1, Median, D_7, P_{80} from graph and Verify your results.

Problem 2. The following tables gives the distribution of 600 workers of a factory:

Weekly wages (in Tk.)	No. of Workers	Weekly wages (in Tk.)	No. of Workers
Below 375	69	600—675	58
375—450	167	675—750	24
450—525	207	750—825	10
525—600	65		

(a) Draw an Ogive from the above data and hence obtain the median value. Check it against calculated value.
(b) Obtain the limits of weekly wages of central 50% of workers.
(c) Calculate standard deviation and the coefficient of variation.

Problem 3. Calculate coefficient of variation and Karl Pearson's coefficient of skewness from the data given below:

Marks	No. of Students cf
Less than 20	18
,, ,, 40	40
,, ,, 60	70
,, ,, 80	90
,, ,, 100	100