1. PARTITION VALUES

Let's Study

- Partition values - Quartiles, deciles, percentiles.
- Graphical location of partition values

Let's Recall

Median – If the numbers in a data are arranged in ascending order, the number at the middle position is called the median of the data. It divides the array of numbers in two equal parts that is the number of scores below and above the median is same. If the number of observations are odd then, Median = $\left(\frac{n+1}{2}\right)^{th}$ observation.

where n = number of observation.

If the number of observations are even then,

Median = Mean of $\left(\frac{n}{2}\right)^{th}$ and $\left(\frac{n+2}{2}\right)^{th}$ observations

$\therefore$ Median =

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

Activity: Complete the following table:

No.	Data	Median
1.	32,33,38,40,43	38
2.	32,33,38,40,43,48	☐
3.	61,62,65,66,68	☐

In previous standard, we have learnt Median for a grouped frequency distribution.

When the number of scores in the data is large, it is difficult to arrange them in ascending order. In such cases the data is divided into groups.

Ex.1 The following table shows frequency distribution of marks of 100 students of 10^{th} class which they obtained in an examination. Find the median marks.

Marks in Exam	0-20	20-40	40-60	60-80	80-100
No. of students	4	20	30	40	6

Solution: N = 100, $\frac{N}{2}$ = 50

C.I.	*f*	*c.f.*
0-20	4	4
20-40	20	24
40-60	30	54
60-80	40	94
80-100	6	100

Median class

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

= class containing observation $\leq \frac{N}{2}$

= 50^{th} observation

$\therefore$ Median class is 40 – 60

Here L = 40, h = 20, f = 30, $c.f.$ = 24

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

$$= 40 + \frac{20}{30}(50 - 24)$$

$$= 40 + \frac{2}{3}(26) = 40 + 17.33$$

$$= 57.33$$

Ex.2 Find Median from the following frequency distribution table.

Age less than (yrs.)	10	20	30	40	50	60
No. of persons	3	10	22	40	54	71

Solution:

Age less than (yrs.)	Class	No. of persons (f)	$c.f.$
10	0-10	3	3
20	10-20	10-□ = □	10
30	□	□-□=□	22
40	□	□-□=□	40
50	□	□-□=□	54
60	50-60	71-54=□	71

Here N = 71, $\frac{N}{\square}$ = 35.5 and h = □

Median class is 30-40

L = □, f = □, $c.f.$ = □

Median = □

$$= 30 + \frac{10}{18}(35.5 - 22)$$

$$= 30 + \square$$

$$= \square$$

Median age of person = 37.5 years.

Introduction –

A certain school is interested in making four divisions of equal strength of 200 students, subject to their marks in their previous class.

The school is interested in a selection criterion for making divisions. Which selection procedure can be more appropriate? Can we select students at random? Will this be fair? Obviously not! Therefore, the school decides to arrange the students in descending order as per their marks and carries out scrutiny for four divisions. This procedure divides students in 4 equal parts of 25% each. In general it may be necessary to divide the data into a required number of equal parts depending upon the situation. This procedure of dividing the data into equal parts is called partitioning and the values that divide the data into the required number of equal parts are called partition values.

Let's Learn

1.1 Partition Values:-

Suppose we have the following data of marks of 15 students in Mathematics.

23,21,26,27,20,18,22,13,14,25,34,28,35,40,38.

Arrange the observations in ascending order of their magnitude.

13,14,18,**20**,21,22,23,**25**,26,27,28,**34**,35,38,40.

What can be said about positions of observations 20, 25, 34? There are 3 observations below 20, 3 observations between 20 and 25, 3 observations between 25 and 34 and 3 observations above 34. These values divide the data in 4 equal parts, so they are called partition values.

Definition : When the observations are arranged in ascending order of their magnitude, and further, they are divided into 'n' equal parts (n=2,3,…), then observations which lie at partitioning positions are called partition values.

We have already studied the concept of one of the partition values namely median. Median

divides the data in two equal parts. Median is a special type of partition value because it divides the data in 2 equal parts. The median, quartiles, deciles and percentiles etc. are termed as partition values, since they divide the given data in equal parts.

We are going to study 3 types of partition values namely quartiles, deciles and percentiles.

Let's Learn

1.2 Quartiles:

Quartiles divide the observations in 4 equal parts when arranged in ascending order of magnitude. There are three quartiles, namely Q_1,Q_2 and Q_3.

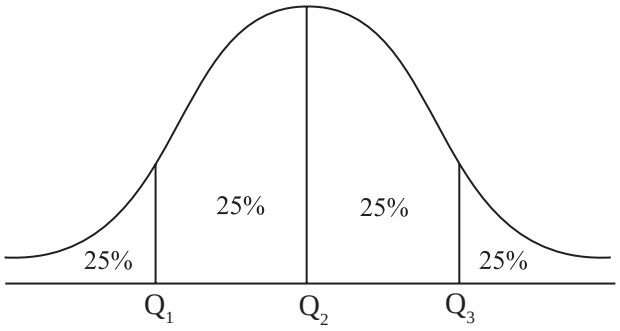


Fig. 1.1

Q_1 is the first or lower quartile. 25% observations are below Q_1 and 75% are above Q_1.

Q_2 is the second quartile (median) which divides data into two equal parts 50% observations are above and 50% below it

Q_3 is the third or upper quartile. 75% of the observations are below Q_3 and 25% are above Q_3.

The formula for quartiles for raw data is as follows:

i^{th} Quartile $= Q_i =$ value of $\left[i\left(\frac{n+1}{4}\right)\right]^{th}$ observation, $i = 1, 2, 3$

where n is the total number of observations.

In particular, if the value of $i\left(\frac{n+1}{4}\right)$ is 6.25, the value of i^{th} quartile is calculated as follows:

i^{th} Quartile = value of 6^{th} observation + 0.25 (value of 7^{th} observation – value of 6^{th} observation)

The formula of quartiles for grouped data is as follows:

$$Q_i = L + \frac{h}{f}\left(\frac{iN}{4} - c.f.\right), \quad i = 1, 2, 3$$

i^{th} Quartile class is the class in which $\left(\frac{iN}{4}\right)^{th}$ observation lies.

Where L = lower boundary of i^{th} quartile class

h = class width of i^{th} quartile class

f = frequency of i^{th} quartile class

$c.f.$ = less than cumulative frequency of the class just preceding i^{th} quartile class

N = total frequency.

SOLVED EXAMPLES

Ex.1: The marks of 19 students are given below:
41, 21, 38, 27,31,45,23,26,29,30,28,25,35, 42,47,50,29,31,35.

Calculate all the quartiles for the above data.

Solution: First arrange the data in ascending order as follows:

21, 23, 25, 26, 27, 28, 29, 29, 30, 31, 31, 35, 35, 38, 41, 42, 45, 47, 50.

Here, n = 19

Q_1 = value of $\left(\frac{19+1}{4}\right)^{th}$ observation

= value of 5^{th} observation

$Q_1 = 27$

$Q_2 = \text{value of } \left[2\left(\frac{19+1}{4}\right)\right]^{th} \text{ observation}$

$= \text{value of } 10^{th} \text{ observation}$

$Q_2 = \square$

$Q_3 = \text{value of } \left[3\left(\frac{19+1}{4}\right)\right]^{th} \text{ observation}$

$= \text{value of } 15^{th} \text{ observation}$

$Q_3 = 41$

Ex.2: Calculate the quartiles for daily wages (₹) by 12 workers: 200,280,310,180,190,170, 320,330,220,210,380,400.

Solution: First arrange the data in ascending order as follows:

170,180,190,200,210,220,280,310,320,330, 380,400

Here, n = 12

$Q_1 = \text{value of } \left(\frac{12+1}{4}\right)^{th} \text{ observation}$

$= \text{value of } 3.25^{th} \text{ observation}$

$= \text{value of } 3^{rd} \text{ observation} + 0.25 \text{ (value of } 4^{th} \text{ observation} - \text{Value of } 3^{rd} \text{ observation)}$

$= 190 + 0.25(200-190)$

$Q_1 = 192.5.$

$Q_2 = \text{value of } \left[2\left(\frac{12+1}{4}\right)\right]^{th} \text{ observation}$

$= \text{value of } 6.5^{th} \text{ observation}$

$= \text{value of } 6^{th} \text{ observation} + 0.5 \text{ (value of } 7^{th} \text{ observation} - \text{value of } 6^{th} \text{ observation)}$

$= 220 + 0.5(280-220)$

$= 220 + 0.5(60)$

$Q_2 = 250$

$Q_3 = \text{value of } \left[3\left(\frac{12+1}{4}\right)\right]^{th} \text{ observation}$

$= \text{value of } 9.75^{th} \text{ observation}$

$= \text{value of } 9^{th} \text{ observation} + 0.75 \text{ (value of } 10^{th} \text{ observation} - \text{value of } 9^{th} \text{ observation)}$

$= 320+0.75(330-320)$

$= 327.5$

Ex.3: Calculate the quartiles for the following data:

Height (in inches)	4.8	4.9	5.0	5.1	5.2	5.3	5.4	5.5	5.6	5.7
No. of Students	5	7	13	16	25	14	9	6	3	2

Solution: Construct the less than cumulative frequency table as follows:

Height (in inches)	No. of students (f)	Less than cumulative frequency
4.8	5	5
4.9	7	12
5.0	13	25
5.1	16	$41 \leftarrow Q_1$
5.2	25	$66 \leftarrow Q_2$
5.3	14	$80 \leftarrow Q_3$
5.4	9	89
5.5	6	95
5.6	3	98
5.7	2	100
Total	100	-

Here, n = 100,

By comparing $\left[i\left(\frac{n+1}{4}\right)\right]^{th}$ with cumulative frequencies, one can easily locate the quartiles.

Q_1 = value of $\left[1\left(\frac{100+1}{4}\right)\right]^{th}$ observation

= value of 25.25^{th} observation

Q_1 = 25^{th} observation + 0.25 (26^{th} observation – 25^{th} observation)

= 5 + 0.25 (5.1−5)

= 5 +0.025

Q_1 = 5.025

Q_2 = value of $\left[2\left(\frac{100+1}{4}\right)\right]^{th}$ observation

= value of 50.5^{th} observation

Q_2 = 5.2

Q_3 = value of $\left[3\left(\frac{100+1}{4}\right)\right]^{th}$ observation

= value of 75.75^{th} observation

Q_3 = 5.3

Ex.4: A highway police department conducted a survey and clocked the speeds of number of cars on a highway. The following distribution was obtained:

Speed below (in kms/hour)	Number of cars
65	19
70	44
75	99
80	184
85	194
90	200

Compute the speed (in kms / hour) below which 75% cars have their speed.

Solution: First, construct classes and frequency table as follows:

Speed (in km / hour)	Number of cars	Less than cumulative frequency
Below 65	19	19
65-70	44-□ = 34	44
□-□	□-□= 55	99
□-□	□-□=□	184 ← Q_3
80-85	□-□=10	194
85-90	□-□=□	200

Compute the speed in km/hour below which 75% cars have their speed that is, we have to calculate the value of Q_3.

Here N = $\sum f$ = □

$\left(\frac{3N}{4}\right)$ = 150

∴ Q_3 lies in the class 75-80

L = □, *c.f.* = 99, h = 5, *f* = □

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

$Q_3 = 75 + \frac{5}{85}(□ - 99)$

$Q_3 = 75 + \frac{255}{85}$

Q_3 = 75+3

Q_3 = □

∴ There are 75% cars passing the highway with speed less than 78 km per hour.

Note : 1) For partition values it is not mandatory to convert discontinuous classes into continuous form. The answer may differ by some decimal figure, which is still correct.

2) If the missing frequency value is in decimal, then approximate it to nearest whole number.

Ex.5: For the following frequency distribution value of Q_2 is 22, find the missing frequency.

Class	Frequency
0.5-9.5	5
10.5-19.5	8
20.5-29.5	?
30.5-39.5	4
40.5-49.5	3

Solution: Let us make the classes continuous and assume that x is a missing frequency. Less than cumulative frequency table is given as follows:

Class	Frequency	Less than cumulative frequency
0-10	5	5
☐–☐	8	☐
☐–☐	x	$13+x \leftarrow Q_2$
30-40	4	☐
☐–☐	3	☐

Here, 2nd quartile (Q_2) = 22

Therefore 20-30 is the second quartile class, Q_2 is given by

$$Q_2 = L + \frac{h}{f}\left(\frac{N}{2} - c.f.\right)$$

L = ☐, $c.f.$ = ☐, h = ☐, $f = x$,

$N = 20 + x$

$$Q_2 = 20 + \frac{10}{x}\left(\frac{20+x}{2} - 13\right)$$

$$Q_2 = 20 + \frac{10}{x}\left(10 + \frac{x}{2} - 13\right)$$

$$Q_2 = 20 + \frac{10}{x}\left(\frac{x}{2} - 3\right)$$

$$Q_2 = 20 + 5 - \frac{30}{x}$$

$$22 = 20 + 5 - \frac{30}{x}$$

$$-3 = -\frac{30}{x}$$

$$x = 10$$

Therefore, the missing frequency is 10.

Ex.6: The following is the distribution of monthly sales of 50 shops on a certain street in a city:

Sales in ₹ ('000)	No. of shops
0-10	3
10-20	?
20-30	20
30-40	12
40-50	?

If the value of Q_1 is 20.75 and two frequencies are missing, find the missing frequencies.

Solution: Here, assume that a and b are missing frequencies. Less than cumulative frequency table is given as follows:

Sales in ₹ ('000)	No. of shops	Less than cumulative frequency
0-10	3	3
10-20	a	$3+a$
20-30	20	$23+a \leftarrow Q_1$
30-40	12	$35+a$
40-50	b	$35+a+b$
Total	$35+a+b$	-

Since total frequency = 50

i.e. $a + b + 35 = 50$

$a + b = 15$ .....(1)

We have given that $Q_1 = 20.75$

20-30 is the first quartile class

Q_1 is given by

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

$L = 20$, $c.f. = 3+a$, $h = 10$, $f = 20$, $N = 50$

$$Q_1 = 20 + \frac{10}{20}\left(\frac{50}{4} - 3 - a\right)$$

$$20.75 = 20 + \frac{1}{2}(12.5 - 3 - a)$$

$$20.75 - 20 = \frac{1}{2}(9.5 - a)$$

$$0.75 = \frac{1}{2}(9.5 - a)$$

$$1.5 = 9.5 - a$$

$$a = 9.5 - 1.5$$

$$a = 8$$

By putting the value of 'a' in equation (1), we get, b = 7

The missing frequencies are 8 and 7 respectively.

EXERCISE 1.1

1. Compute all the quartiles for the following series of observations :

   16, 14.9, 11.5, 11.8, 11.1, 14.5, 14, 12, 10.9, 10.7, 10.6, 10.5, 13.5, 13, 12.6.

2. The heights (in cm.) of 10 students are given below :

   148, 171, 158, 151, 154, 159, 152, 163, 171, 145.

   Calculate Q_1 and Q_3 for the above data.

3. Monthly consumption of electricity (in units) of families in a certain locality is given below :

   205, 201, 190, 188, 194, 172, 210, 225, 215, 232, 260, 230.

   Calculate electricity consumption (in units) below which 25% of families lie.

4. For the following data of daily expenditure of families (in ₹), compute the expenditure below which 75% of families include their expenditure.

Daily Expenditure (in ₹)	350	450	550	650	750
No. of families	16	19	24	28	13

5. Calculate all the quartiles for the following frequency distribution :

No. of E-transactions per day	0	1	2	3	4	5	6	7
No. of days	10	35	45	95	64	32	10	9

6. The following is the frequency distribution of heights of 200 male adults in a factory :

Height in cm.	No. of male adults
145-150	4
150-155	6
155-160	25
160-165	57
165-170	64
170-175	30
175-180	8
180-185	6

   Find the central height.

7. The following is the data of pocket expenditure per week of 50 students in a class. It is known that the median of the distribution is ₹120. Find the missing frequencies.

Expenditure per week (in ₹)	0-50	50-100	100-150	150-200	200-250
No. of students	7	?	15	?	3

8. The following is the distribution of 160 workers according to the wages in a certain factory :

Wages more than (in ₹)	No. of workers
8000	160
9000	155
10000	137
11000	103
12000	57
13000	23
14000	10
15000	1
16000	0

Determine the values of all quartiles and interpret the results.

9. Following is grouped data for duration of fixed deposits of 100 senior citizens from a certain bank :

Fixed deposits (in days)	0-180	180-360	360-540	540-720	720-900
No. of senior citizens	15	20	25	30	10

Calculate the limits of fixed deposits of central 50% senior citizens.

10. Find the missing frequency given that the median of distribution is 1504.

Life in hours	950-1150	1150-1350	1350-1550	1550-1750	1750-1950	1950-2150
No. of bulbs	20	43	100	-	23	13

1.3 Deciles and Percentiles:

Let's Learn

Deciles:

Deciles are values, which divide the observations in 10 equal parts when arranged in ascending order of magnitude

Deciles are denoted by D_1, D_2, …. D_9

The formula of deciles for raw data is given below:

i^{th} Decile $= D_i$

$= $ Value of observation, $\left[i\left(\frac{n+1}{10}\right)\right]^{th}$

$i = 1, 2, \ldots, 9$

Note:- If the above value is not an integer then use the method explained in calculation of Quartiles. (Refer page no. 3).

The formula of deciles for grouped data is,

$$D_i = L + \frac{h}{f}\left(\frac{iN}{100} - c.f.\right), \qquad i = 1, 2, \ldots, 9$$

Here L = lower boundary of i^{th} decile class

h = class width of i^{th} decile class

f = frequency of decile class

$c.f.$ = less than cumulative frequency of the class preceding i^{th} decile class

N = total frequency

Let's Learn

Percentiles:

Percentiles are values which divide observations in 100 equal parts when they are arranged in ascending order of magnitude. Percentiles are denoted by P_1, P_2,….,P_{99}.

The formula of percentiles for raw data is given below:

i^{th} Percentile $= P_i$

$= $ Value of observation, $\left[i\left(\frac{n+1}{100}\right)\right]^{th}$

$i = 1, 2, \ldots, 99.$

Note:- If the above value is not an integer then use the method explained in calculation of Quartiles. (Refer page no. 3).

The formula of percentile for grouped data is,

$$P_i = L + \frac{h}{f}\left(\frac{iN}{100} - c.f.\right),\ i = 1, 2,\ldots,99$$

Here, L = lower boundary of i^{th} percentile class

h = class width of i^{th} percentile class

f = frequency of i^{th} percentile class

$c.f.$ = less than cumulative frequency of the class just preceding i^{th} percentile class

N = total frequency

Let's Learn

Relations among quartiles, deciles and percentiles:

From definition it can be seen that,

(1) $Q_2 = D_5 = P_{50}$

(2) $Q_1 = P_{25}$

(3) $Q_3 = P_{75}$

Let's Think

Can you find $D_1 = P_{\square}$, $P_{30} = D_{\square}$, $D_{\square} = P_{70}$?

SOLVED EXAMPLES

Ex.1: Calculate 3rd decile and 70th percentile for the following data:

841, 289, 325, 225, 784, 729, 625, 400, 324, 169.

Solution: First, we arrange the data in ascending order as follows:

169, 225, 289, 324, 325, 400, 625, 729, 784, 841

Here, $n = 10$

D_3 = Value of $\left[3\left(\frac{10+1}{10}\right)\right]^{th}$ observation

= value of 3.3^{rd} observation

= value of 3^{rd} observation + 0.3

(value of 4^{th} observation – value of 3^{rd} observation)

= 289 +0.3 (324 – 289)

D_3 = 299.5

P_{70} = Value of $\left[70\left(\frac{10+1}{100}\right)\right]^{th}$ observation

= value of 7.7^{th} observation

= value of 7^{th} observation + 0.7

(value of 8^{th} observation – value of 7^{th} observation)

= 625 +0.7 (729– 625)

= 625+72.8

P_{70} = 697.8

Ex.2: Find D_4 and P_{55} for the following data:

No.of defective Products	30	35	40	45	50	55	60
No. of firms	12	35	10	15	8	7	8

Solution: Prepare less than cumulative frequency table as below:

No. of defective products	No. of firms	Less than cumulative frequency
30	12	12
35	35	47 ← D_4
40	10	57 ← P_{55}
45	15	72

50	8	80
55	7	87
60	8	95

n = 95

D_4 = Value of $\left[4\left(\frac{95+1}{10}\right)\right]^{th}$ observation

= value of 38.4^{th} observation

D_4 = 35

P_{55} = Value of $\left[55\left(\frac{95+1}{100}\right)\right]^{th}$ observation

= value of 52.8^{th} observation

P_{55} = 40

Ex.3: Calculate the 4th decile and 21st percentile from the following data:

Profit (in lakh ₹)	0.5-4.5	5.5-9.5	10.5-14.5	15.5-19.5	20.5-24.5
No. of firms	7	18	25	30	20

Solution: Make the classes continuous and construct less than cumulative frequency table as follows:

Profit (in lakh ₹)	Number of firms (f)	Less than cumulative frequency
0-5	7	7
□–□	18	$25 \leftarrow P_{21}$
□–□	25	$50 \leftarrow D_4$
15-20	30	□
□–□	20	□

D_4 = Value of $\left[4\left(\frac{100}{10}\right)\right]^{th}$ observation

D_4 = value of 40^{th} observation

D_4 lies in the class 10-15

L = lower boundary of fourth decile class = □

h = class width of fourth decile class = □

f = frequency of fourth decile class = □

$c.f.$ = less than cumulative frequency of the class preceding fourth decile class = □

N = total frequency = □

$D_4 = L + \frac{h}{f}\left(\frac{4N}{10} - c.f.\right)$

$= 10 + \frac{5}{25}\left(\frac{4\times100}{10} - 25\right)$

$= 10 + \frac{5}{25}(40-25)$

= 10+3

D_4 = 13

For percentile

P_{21} = Value of $\left(\frac{21\times100}{100}\right)^{th}$ observation

= value of 21^{st} observation

P_{21} lies in the class 5-10

L = lower boundary of 21^{st} percentile class = □

h = class width of 21^{st} percentile class = □

f = frequency of 21^{st} percentile class = □

$c.f.$ = less than cumulative frequency of the class preceding 21^{st} percentile = □

N = total frequency = 100

$P_{21} = L + \frac{h}{f}\left(\frac{21N}{100} - c.f.\right)$

$= 5 + \frac{5}{18}\left(\frac{21\times100}{100} - □\right)$

$$= 5 + \frac{5}{18}(21-7)$$

$$= 5 + \frac{70}{18}$$

$$= 5 + 3.89$$

$$P_{21} = 8.89$$

Ex.4: A study related to the journey of office files through various departments was conducted. The distribution of time (in months) taken by a file to come back to the initiating department is given below:

Time	1-2	2-3	3-4	4-5	5-6	6 and above
No. of files	30	45	40	50	6	4

Find: (a) the time by which 75% of the files come back.

(b) the time by which at most 60% of the files come back.

Solution: Construct less than cumulative frequency table as follows:

Time (in months)	Number of files (f)	Less than cumulative frequency
1-2	30	☐
2-3	45	☐
3-4	40	$115 \leftarrow P_{60}$
4-5	50	$165 \leftarrow P_{75}$
5-6	6	☐
6 and above	4	☐

To find the time by which 75% of the files come back i.e. we have to find the value of P_{75}

$$P_{75} = \text{value of } \left(\frac{75\times175}{100}\right)^{th} \text{ observation}$$

$$= \text{value of } 131.25^{th} \text{ observation}$$

P_{75} lies in the class 4-5

L = lower boundary of 75^{th} percentile class = ☐

h = class width of 75^{th} percentile class = ☐

f = frequency of 75^{th} percentile class = ☐

$c.f.$ = less than cumulative frequency of the class preceding 75^{th} percentile = ☐

N = total frequency = ☐

$$P_{75} = L + \frac{h}{f}\left(\frac{75N}{100} - c.f.\right)$$

$$= \square + \frac{1}{50}\left(\frac{75\times175}{100} - \square\right)$$

$$= 4 + \frac{1}{50}(131.25-115)$$

$$= 4 + \frac{16.25}{50}$$

$$= 4 + 0.325$$

$$P_{75} = 4.325$$

Thus the time by which 75% of the files come back is 4.325 months.

We have to find the time by which at most 60% of the files come back. that is we have to find P_{60}

For 60^{th} percentile

$$P_{60} = \text{value of } \left(\frac{60\times175}{100}\right)^{th} \text{ observation}$$

P_{60} = value of 105^{th} observation

P_{60} lies in the class 3-4

L = lower boundary of 60^{th} percentile class = ☐

h = class width of 60^{th} percentile class = ☐

f = frequency of 60^{th} percentile class = ☐

$c.f.$ = less than cumulative frequency of the class preceding 60^{th} percentile = $\square$

N = total frequency = 175

$$P_{60} = L + \frac{h}{f}\left(\frac{60N}{100} - c.f.\right)$$

$$= \square + \frac{1}{40}\left(\frac{60\times175}{100} - 75\right)$$

$$= 3 + \frac{1}{40}(105 - 75)$$

$$= 3 + \frac{30}{40}$$

$$= 3+0.75$$

$$P_{60} = 3.75$$

Thus the time by which at most 60% of the files come back is 3.75 months.

Ex.5: Following is the frequency distribution of number of days when students were absent in a particular class:

Number of days absent	No. of students
Below 10	15
Below 20	35
Below 30	60
Below 40	84
Below 50	106
Below 60	120
Below 70	125

Calculate D_6 and P_{30} for the above data.

Solution: Convert the data in groups and construct less than cumulative frequency table as follows:

Number of days absent	Number of students (f)	Less than cumulative frequency
$\square$–$\square$	15	15
10-20	$\square$	35
$\square$–$\square$	25	$60 \leftarrow P_{30}$
30-40	$\square$	$84 \leftarrow D_6$
$\square$–$\square$	22	106
50-60	$\square$	120
$\square$–$\square$	5	125

D_6 = value of $\left(\frac{6\times125}{10}\right)^{th}$ observation

= value of 75^{th} observation

D_6 lies in the class 30-40

L = lower boundary of sixth decile class = $\square$

h = class width of sixth decile class = $\square$

f = frequency of sixth decile class = $\square$

$c.f.$ = less than cumulative frequency of the class preceding 6^{th} decile = $\square$

N = total frequency = $\square$

$$D_6 = L + \frac{h}{f}\left(\frac{6N}{10} - c.f.\right)$$

$$= \square + \frac{10}{24}\left(\frac{6\times125}{10} - 60\right)$$

$$= 30 + \frac{10}{24}(75 - 60)$$

$$= 30 + \frac{150}{24}$$

$$= 30+6.25$$

$$D_6 = 36.25$$

i.e. 60% students are absent for less than 36 days.

$$P_{30} = \text{value of } \left(\frac{30\times125}{100}\right)^{th} \text{ observation}$$

P_{30} = value of 37.5^{th} observation

P_{30} lies in the class $\square$

L = lower boundary of 30^{th} percentile class = 20

h = class width of 30^{th} percentile class = 10

f = frequency of 30^{th} percentile class = 25

$c.f.$ = less than cumulative frequency of the class preceding 30^{th} percentile class = 35

N = total frequency = 125

$$P_{30} = L + \frac{h}{f}\left(\frac{30N}{100} - c.f.\right)$$

$$= \square + \frac{10}{25}\left(\frac{30\times125}{100} - \square\right)$$

$$= 20 + \frac{10}{25}(37.5-35)$$

$$= 20 + \frac{10\times2.5}{25}$$

$$= 20+1$$

$$P_{30} = 21$$

i.e. 30% students are absent for less than 21 days.

Ex.6: Following is the distribution of marks in Statistics obtained by 50 students

Marks	No. of students
More than 0	50
More than 10	46
More than 20	40
More than 30	20
More than 40	10
More than 50	3

If 60% students passed the examination, find the minimum marks required to pass examination.

Solution: Construct the frequency and less than cumulative frequency table as follows:

Marks	Number of students (f)	Less than cumulative frequency
$\square$–$\square$	4	4
$\square$–$\square$	$\square$	10
20-30	30-10=20	$30 \leftarrow P_{40}$
$\square$–$\square$	$\square$	40
$\square$–$\square$	$\square$	47
50-60	$\square$	50

In order to find the minimum marks obtained by a passed student if 60% students passed, 40%will fail in the examination. So minimum marks required for passing is to cross P_{40}.

P_{40} = value of $\left(\frac{40\times50}{100}\right)^{th}$ observation

= value of 20^{th} observation

P_{40} lies in the class $\square$

L = lower boundary of 40^{th} percentile class = $\square$

h = class width of 40^{th} percentile class = $\square$

f = frequency of 40^{th} percentile class = 20

$c.f.$ = less than cumulative frequency of the class preceding 40^{th} percentile = $\square$

N = total frequency = $\square$

$$= L + \frac{h}{f}\left(\frac{40N}{100} - c.f.\right)$$

$$= \square + \frac{10}{20}\left(\frac{40\times50}{100} - 10\right)$$

$$= 20 + \frac{10}{20}(20-10)$$

$$= 20 + \frac{100}{20}$$

$= 20 + 5$

$P_{40} = 25$

Hence if 60% students pass the test, the minimum marks for passing of a candidate should be 25.

Ex.7: Following is the distribution of age of 500 workers, find the percentage of workers whose age is more than 45 years.

Age	20-30	30-40	40-50	50-60
No. of workers	80	160	180	80

Solution: construct the frequency and less than cumulative frequency table

Class Interval	Frequency (f)	Less than cumulative frequency ($l.c.f.$)
20-30	80	☐
30-40	160	☐
40-50	180	$420 \leftarrow P_x$
50-60	80	500

Since, $P_x = 45$ lies in the class 40-50

L = lower boundary of P_x class = 40

h = class width of P_x class = 10

f = frequency of P_x class = 180

$c.f.$ = less than cumulative frequency of the class preceding P_x class = 240

N = total frequency = 500

$$P_x = L + \frac{h}{f}\left(\frac{xN}{100} - c.f.\right)$$

$$45 = 40 + \frac{10}{180}\left(\frac{x \times 500}{100} - 240\right)$$

$$45 - 40 = \frac{10}{180}\left(\frac{x \times 500}{100} - 240\right)$$

$$\frac{180 \times 5}{10} = \left(\frac{x \times 500}{100} - 240\right)$$

$$90+240 = 5x$$

$$330 = 5x$$

$$x = 66$$

66% workers have age below 45 years and 34% workers have age more than 45 years.

Alternative Method:

Above example can also be solved by the following method. Since the percentage of workers having age more than 45 years is required, we find the class in which '45' lies. Clearly '45' lies in the class 40-50. The class 40-50 is divided by 45 as 40-45 and 45-50. The ratio of width of these subclasses is 5: 5, i.e. 1: 1. Hence, the frequency of the class 40-50 which is 180 should be divided in the ratio 1: 1. So the number of workers in both the subclasses is 90.

Hence, the number of workers having age above 45 is 90+80=170.

Their required percentage = $\frac{170}{500} \times 100 = 34$.

34% workers have age more than 45 years.

Ex.8: If the 2nd decile of the distribution given below is 13.75, find the missing value.

Mid points	5	15	25	35	45
Frequency	2	?	10	3	2

Solution: construct the frequency and less than cumulative frequency table

Class Interval	Frequency f	Less than cumulative frequency
0-10	2	2
10-20	x	$2 + x \leftarrow D_2$

20-30	10	12 + x
30-40	3	15 + x
40-50	2	17 + x

D_2 is 13.75 lies in the class 10-20

L = lower boundary of 2nd decile class = 10

h = class width of 2nd decile class = 10

f = frequency of 2nd decile class = x

c.f. = less than cumulative frequency of the class preceding 2nd decile class = 2

N = total frequency = 17 + x

$$D_2 = L + \frac{h}{f}\left(\frac{2N}{10} - c.f.\right)$$

$$= 10 + \frac{10}{x}\left(\frac{2(17+x)}{10} - 2\right)$$

$$13.75 = 10 + \frac{10}{x}\left(\frac{34+2x}{10} - 2\right)$$

$$13.75\text{-}10 = \frac{10}{x}\left(\frac{34+2x}{10} - 2\right)$$

$$\frac{3.75x}{10} + 2 = \frac{34+2x}{10}$$

$3.75x + 20 = 34 + 2x$

$1.75x = 14$

$x = 8$

The missing value is 8.

EXERCISE 1.2

1. Calculate D_6 and P_{85} for the following data : 79,82,36,38,51,72,68,70,64,63

2. The daily wages (in ₹) of 15 laboures are as follows :

   230,400,350,200,250,380,210,225,375,180, 375,450,300,350,250

   Calculate D_8 and P_{90}

3. Calculate 2nd decile and 65th percentile for the following :

x	80	100	120	145	200	280	310	380	400	410
f	15	18	25	27	40	25	19	16	8	7

4. From the following data calculate the rent of 15th, 65th & 91st house.

House Rent (in ₹)	11000	12000	13000	15000	14000	16000	17000	18000
No. of houses	25	17	13	14	15	8	6	2

5. The following frequency distribution shows the weight of students in a class.

Weight (in Kg)	40	45	50	55	60	65
Number of Students	15	40	29	21	10	5

a) Find the percentage of students whose weight is more than 50 kg.

b) If the weight column provided is of mid values then find the percentage of students whose weight is more than 50 kg.

6. Calculate D_4 and P_{48} from the following data:

Mid Value	2.5	7.5	12.5	17.5	22.5	Total
Fre-quency	7	18	25	30	20	100

7. Calculate D_9 and P_{20} of the following distribution.

Length (in Inches)	0-20	20-40	40-60	60-80	80-100	100-120
No. of units	1	14	35	85	90	15

8. Weekly wages for group of 100 persons are given below :

Wages (in ₹)	0-500	500-1000	1000-1500	1500-2000	2000-2500
No. of persons	7	?	25	30	?

D_3 for this group is ₹ 1100 Calculate the missing frequencies.

9. The weekly profit (in rupees) of 100 shops are distributed as follows :

Profit per shop	No. of Shops
0-1000	10
1000-2000	16
2000-3000	26
3000-4000	20
4000-5000	20
5000-6000	5
6000-7000	3

Find the limits of the profit of middle 60% of the shops.

10. In a particular factory, workers produce various types of output units

The following distribution was obtained

Outputs units Produced	No. of workers
70-74	40
75-79	45
80-84	50
85-89	60
90-94	70
95-99	80
100-104	100

Find the percentage of workers who have produced less than 82 output units.

Let's Learn

1.4 Graphical location of partition values:

We have already studied how to determine partition values by numerical methods. Now, let us study a graphical method to determine partition values of grouped frequency distributions.

(i) Ogive is a cumulative frequency curve. Since there are two types of cumulative frequencies, there are two types of ogives, namely less than ogive (less than type cumulative curve) and more than ogive (more than type cumulative curve).

(a) To draw less than ogive :

Step 1 : Compute less than cumulative frequencies (*lcf*) for the given frequency distribution. Assign zero *lcf* for the preceding class of the first class

Step 2 : Take upper class boundaries on the X-axis and *lcf.* on the Y-axis in XY-plane

Step 3 : Plot the points (x,y) where x = upper class boundary, $y = l.c.f$, for all classes.

Step 4 : Join these points by a smooth free hand curve. This curve is called "less than ogive".

(b) To draw more than ogive :

Step 1 : Compute more than cumulative frequencies (*mcf*) for the given frequency distribution. Assign zero mcf for the succeeding class of the last class.

Step 2 : Take lower boundaries on the X-axis and mcf on the Y-axis in XY-plane.

Step 3 : Plot the points (x,y), (x = lower class boundary, $y = mcf$) for all classes.

Step 4 : Join these points by a smooth free hand. This is called more than ogive.

(ii) Locate $\frac{N}{2}$ on the Y-axis and draw a line parallel to X-axis from this point, meeting the ogive at a point say P. Draw a perpendicular to X-axis from P. Median is the value at the foot of perpendicular.

Similarly, other partition values can also be located.

i^{th} quartile (Q_i) is determined by drawing a line parallel to X-axis from y = $i\,\frac{N}{4}$

i^{th} decile (D_i) is obtained by drawing a line parallel to X-axis from y = $i\,\frac{N}{10}$

i^{th} percentile (P_i) is obtained by drawing line parallel to X-axis from y = $i\,\frac{N}{100}$.

Remarks:

(1) An alternative method of determining median by graphical method is as follows:

Draw both types of ogives on the same graph paper. The perpendicular to X-axis drawn from the point of intersection of two ogives meets it at the median.

(2) The graphical method may not give exact partition values. The values obtained by graphical method are approximate.

SOLVED EXAMPLES

Ex. 1: For the following data of marks in Statistics, determine, Q_2, D_8 and P_{60} graphically.

Marks	0-10	10-20	20-30	30-40	40-50	50-60
Fre-quency	4	6	20	10	7	3

Solution:

Marks	Frequency	Less than cumulative-Frequency
0-10	4	4
10-20	6	10
20-30	20	30
30-40	10	40
40-50	7	47
50-60	3	50

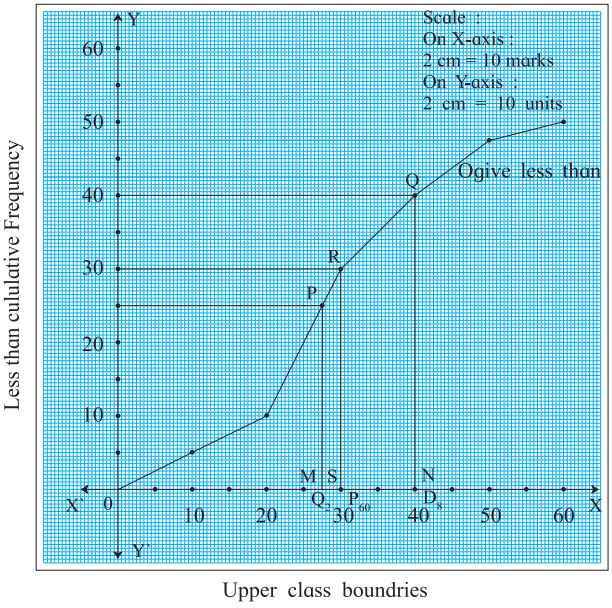


Fig. 1.2

Plotting the less than cumulative frequency against the corresponding upper class boundaries and joining these points by a smooth curve, we get less than ogive as shown in figure 1.2

(i) At the frequency $\frac{N}{2} = \frac{50}{2} = 25$, draw a line parallel to X-axis as shown in fig. 1.2

Then OM = 27.5 is the Q_2 marks.

(ii) At the frequency $\frac{8N}{10} = \frac{8\times 50}{10} = 40$, draw a line parallel to X-axis as shown in fig. 1.2

Then, D_8 = ON = 40

Similarly P_{60} = OS = 30

Ex.2: The following distribution gives the pattern of overtime work per week done by 100 employees of a company. Calculate median, first quartile and seventh decile graphically. Also estimate numbers of employees who worked less than 11 hours per week.

Overtime (in hours)	No. of employees
10-14	11
15-19	20
20-24	35
25-29	20
30-34	8
35-39	6

Solution: Here the class intervals are discontinuous.

Convert the class intervals into continuous classes

Marks	Frequency	Less than cumulative frequency
9.5-14.5	11	11
14.5-19.5	20	31
19.5-24.5	35	66
24.5-29.5	20	86
29.5-34.5	8	94
34.5-39.5	6	100

Plotting upper class boundaries against less than cumulative frequencies, we get ogive as shown in the figure 1.3

To find out median, Q_1 and P_{70} graphically,

draw lines parallel to X-axis from frequencies $\frac{N}{2}$, $\frac{N}{4}$ and $\frac{7N}{10}$ respectively. Draw perpendiculars on X-axis from P, Q and R.

Then Q_1 = OL = 18

Median = OM = 23 approximately

P_{70} = ON = 25 approximately

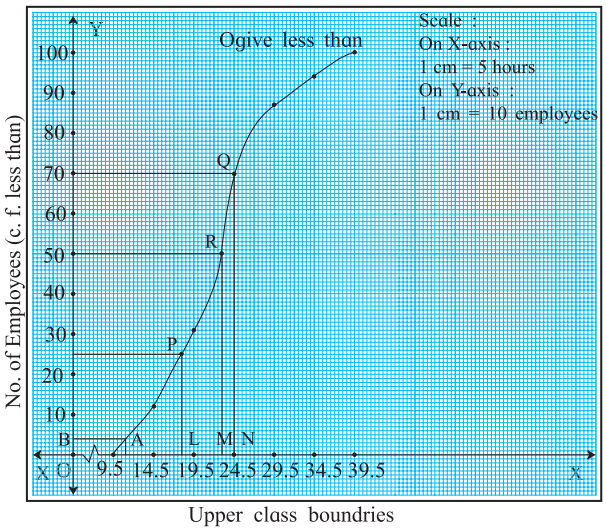


Fig. 1.3

To find number of employees who worked less than 11 hours per week ,draw a line parallel to Y-axis from upper class boundaries equal to 11 which meets the curve in A. Draw perpendicular AB on Y-axis. Read the value at the foot of perpendicular namely B.

The number of workers who worked less than 11 hours = OB = 4

Let's Think

Can you solve above problem by drawing more than ogive?

EXERCISE 1.3

1. The following table gives frequency distribution of marks of 100 students in an examination.

Marks	15-20	20-25	25-30	30-35	35-40	40-45	45-50
No. of students	9	12	23	31	10	8	7

Determine D_6, Q_1 and P_{85} graphically.

2. The following table gives the distribution of daily wages of 500 families in a certain city.

Daily wages	No. of families
Below 100	50
100-200	150
200-300	180
300-400	50
400-500	40
500-600	20
600 above	10

Draw a 'less than' ogive for the above data Determine the median income and obtain the limits of income of central 50% of the families.

3. From the following distribution, determine median graphically.

Daily wages (in ₹)	No. of employees
Above 300	520
Above 400	470
Above 500	399
Above 600	210
Above 700	105
Above 800	45
Above 900	7

4. The following frequency distribution shows the profit (in ₹) of shops in a particular area of city.

Profit per shop (in '000)	No. of shops
0-10	12
10-20	18
20-30	27
30-40	20
40-50	17
50-60	6

Find graphically

(i) The limits of middle 40% shops.

(ii) The number of shops having profit less than 35,000 rupees.

5. The following is frequency distribution of over time (per week) performed by various workers from a certain company.

Determine the values of D_2, Q_2 and P_{61} graphically.

Overtime (in hours)	Below 8	8-12	12-16	16-20	20-24	24 and above
No. of workers	4	8	16	18	20	14

6. Draw ogive for the following data and hence find the values of D_1, Q_1.P_{40}.

Marks less than	10	20	30	40	50	60	70	80	90
No. of students	4	6	24	46	67	86	96	99	100

7. The following table shows the age distribution of head of the families in a certain country. Determine the third, fifth and eighth decile of the distribution graphically.

Age of head of family (in years)	Numbers (million)
Under 35	46
35-45	85
45-55	64
55-65	75
65-75	90
75 & Above	40

8. The following table gives the distribution of females in an Indian village. Determine the median age graphically.

Age group	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No. of females (in '000')	175	100	68	48	25	50	23	8	2	1

9. Draw ogive for the following distribution and hence find graphically the limits of weight of middle 50% fishes.

Weight of fishes (in gms)	800-890	900-990	1000-1090	1100-1190	1200-1290	1300-1390	1400-1490
No. of fishes	8	16	20	25	40	6	5

10. Find graphically the values of D_3 and P_{65} for the data given below :

I.Q. of students	60-69	70-79	80-89	90-99	100-109	110-119	120-129
No. of students	20	40	50	50	20	10	10

Let's Remember

- Partition values divide the given distribution into a number of equal parts.
- Quartiles, deciles and percentiles divide the distribution into 4, 10, and 100 equal parts respectively.
- The formulae for raw data:

i^{th} Quartile = Q_i = value of $\left[i\left(\frac{n+1}{4}\right)\right]^{th}$ observation,

i^{th} Decile = D_i = value of $\left[i\left(\frac{n+1}{10}\right)\right]^{th}$ observation,

i^{th} Percentile = P_i = value of $\left[i\left(\frac{n+1}{100}\right)\right]^{th}$ observation,

- The formulae for grouped continuous data :

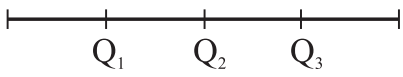
$$Q_i = L + \frac{h}{f}\left(\frac{iN}{4} - c.f.\right),\ i = 1,2,3$$

$$D_i = L + \frac{h}{f}\left(\frac{iN}{10} - c.f.\right),\ i = 1,2,3,.....,9$$

$$P_i = L + \frac{h}{f}\left(\frac{iN}{100} - c.f.\right),\ i = 1,2,.....,99$$

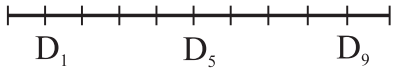
- Approximate partition values can be determined by using ogive.
- Median = $Q_2 = D_5 = P_{50}$

Quartiles



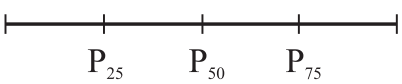
$Q_1 < Q_2 < Q_3$

Deciles



$D_1 < D_2 < D_3 < \ldots\ldots\ldots < D_9$

Percentiles



$P_1 < P_2 < P_3 < \ldots\ldots < P_{99}$.

MISCELLANEOUS EXERCISE

1. The data gives number of accidents per day on a railway track Compute Q_2, P_{17} and D_7.

   4,2,3,5,6,3,4,1,2,3,2,3,4,3,2.

2. The distribution of daily sales of shoes (size wise) for 100 days from a certain shop is :

Size of shoes	2	4	3	5	7	6	8
No. of days	14	20	13	19	13	13	8

   Compute Q_2 , D_1 and P_{95} .

3. Ten students appeared for a test in Mathematics and Statistics and they obtained the marks as follows:

Sr. No.	1	2	3	4	5	6	7	8	9	10
Marks in Mathe-matics	42	38	36	32	23	25	35	37	25	23
Marks in Statistics	22	26	29	34	50	45	23	28	32	36

   If the median will be the criteria, in which. subject, the level of knowledge of the students is higher?

4. In the frequency distribution of families given below, the number of families corresponding to expenditure groups 2000-4000 is missing from the table. However value of 25th percentile is 2880. Find the missing frequency.

Weekly Expenditure (₹ 1000)	0-2	2-4	4-6	6-8	8-10
No. of families	14	?	39	7	15

5. Calculate Q_1, D_6 and P_{15} for the following data :

Mid value	25	75	125	175	225	275
Frequency	10	70	80	100	150	90

6. Daily income for a group of 100 workers are given below :

Daily Income (in ₹)	0-50	50-100	100-150	150-200	200-250
No. of persons	7	?	25	30	?

P_{30} for this group is ₹ 110. Calculate missing frequencies.

7. The distribution of a sample of students appearing for a C.A. examination is:

Marks	0-100	100-200	200-300	300-400	400-500	500-600
No. of students	130	150	190	220	280	130

Help C.A. institute to decide cut off marks for qualifying the examination when 3% students pass the examination.

8. Determine graphically the value of median, D_3 and P_{35} for the data given below :

Class	10-15	15-20	20-25	25-30	30-35	35-40	40-45
Frequency	8	14	8	25	15	14	6

9. The I.Q. test of 500 students of a college is as follows:

I.Q.	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Number of students	41	52	64	180	67	45	40	11

Find graphically the number of students whose I.Q. is more than 55 graphically.

10. Draw an ogive for the following distribution. Determine the median graphically and verify your result by mathematical formula.

Height (in cms.)	No. of students
145-150	2
150-155	5
155-160	9
160-165	15
165-170	16
170-175	7
175-180	5
180-185	1

11 In a group of 25 students 7 students failed and 6 students got distinction and the marks of remaining 12 students are 61,36,44,59, 52,56,41,37,39,38,41,64. Find the median marks of the whole group.

12. The median weight of a group of 79 students is found to be 55 kg 6 more students are added to this group whose weights are 50,51,52,59.5,60,61kg. What will be the value of median of the combined group if the lowest and the highest weights were 53kg and 59 kg respectively?

13. The median of the following incomplete table is 92. Find the missing frequencies:

C I	30-50	50-70	70-90	90-110	110-130	130-150	Total
f	6	?	18	20	?	10	80

14. A company produces tables which are packed in batches of 100. An analysis of the defective tubes in different batches has received the following information:

No. of defective tubes	Less than 5	5-9	10-14	15-19	20-24	25-29	30 and above
No. of tubes	45	51	84	39	20	8	4

Estimate the number of defective tubes in the central batch.

15. In a college there are 500 students in junior college, 5% score less than 25 marks, 68 score from 26 to 30 marks , 30% score from 31 to 35 marks, 70 score from 36 to 40 marks, 20% score from 41 to 45 marks and the rest score 46 and above marks. What is the median marks?

16. Draw a cumulative frequency curve more than type for the following data and hence locate Q_1 and Q_3. Also find the number of workers with daily wages

(i) Between ₹ 170 and ₹ 260

(ii) less than ₹ 260

Daily wages more than (₹)	100	150	200	250	300	350	400	450	500
No. of workers	200	188	160	124	74	49	31	15	5

17. Draw ogive of both the types for the following frequency distribution and hence find median.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No. of students	5	5	8	12	16	15	10	8	5	2

18. Find Q_1, D_6 and P_{78} for the following data :

C.I.	8-8.95	9-9.95	10-10.95	11-11.95	12-12.95
f	5	10	20	10	5

19.

Weight (kg)	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80
No. of persons	4	15	20	30	20	10	8	4

For above data, find all quartiles and number of persons weighing between 57 kg and 72 kg.

20. For the following data showing weights of 100 employees, find the maximum weight of the lightest 25% of employees.

Weight (kg)	45-50	50-55	55-60	60-65	65-70	70-75	75-80
No. of em-ployees	6	8	15	26	20	14	11

Activity 1.1

Divide the number of students in your class according to their marks in Mathematics in 10^{th} standard in

(a) 4 equal parts.

(b) 10 equal parts.

Activity 1.2

Find scores in an examination of students in your class.

(a) The number of students :

(b) The average marks :

(c) The median marks :

Who scored the median marks :

(d) Find the score so that 75% students have scores below it.

Activity 1.3

In a class of 80 students, find number of students with 100% attendance, < 75% attendance, < 50% attendance, and < 25% attendance.

Graphically find the number of students in the black list (that is, students who have < 50% attendance)

Activity 1.4

Think and give answer.

If the building is constructed with 114 floors and we need to design six fire exits equally placed on floors of the building. Suggest the floor numbers on which the fire exits can be constructed.