QUARTILES, DECILES, AND PERCENTILES

Median is the middle point in the axis frequency distribution curve, and it divides the area under the curve into two equal parts, having the same area in the left, and in the right.

On similar basis, the area under the curve may be divided into four equal areas or parts, called as quartiles. In the same procedure divide the area for ten equally pieces and each area is called deciles. Finally where divided the area for hundred equally pieces and each area is called percentiles.

Quartile

* A quartile is a statistical term that describes a division of observations into four defined intervals based on the values of the data and how they compare to the entire set of observations.
* The quartile measures the spread of values above and below the mean by dividing the distribution into four groups.
* A quartile divides data into three points—a lower quartile, median, and upper quartile—to form four groups of the dataset.

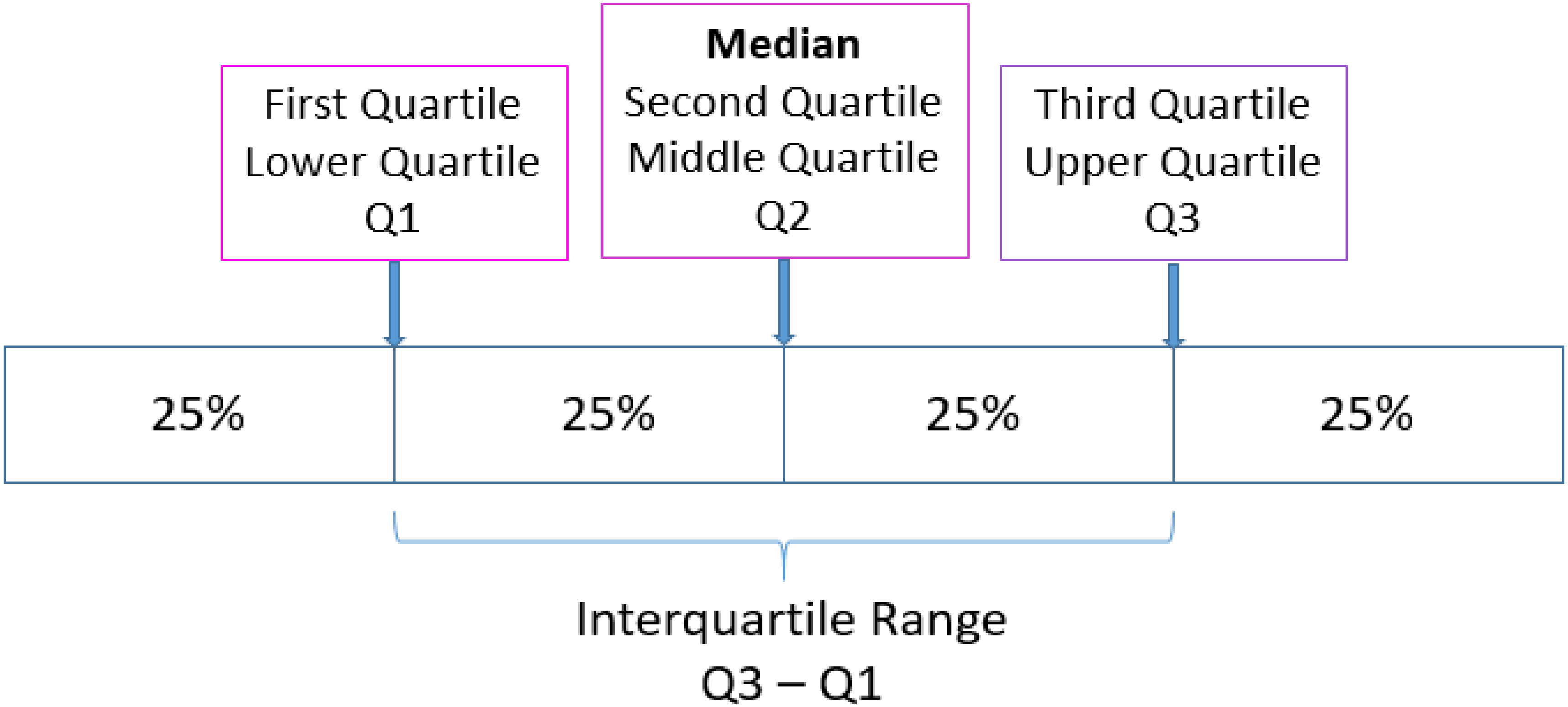
How Quartiles Work

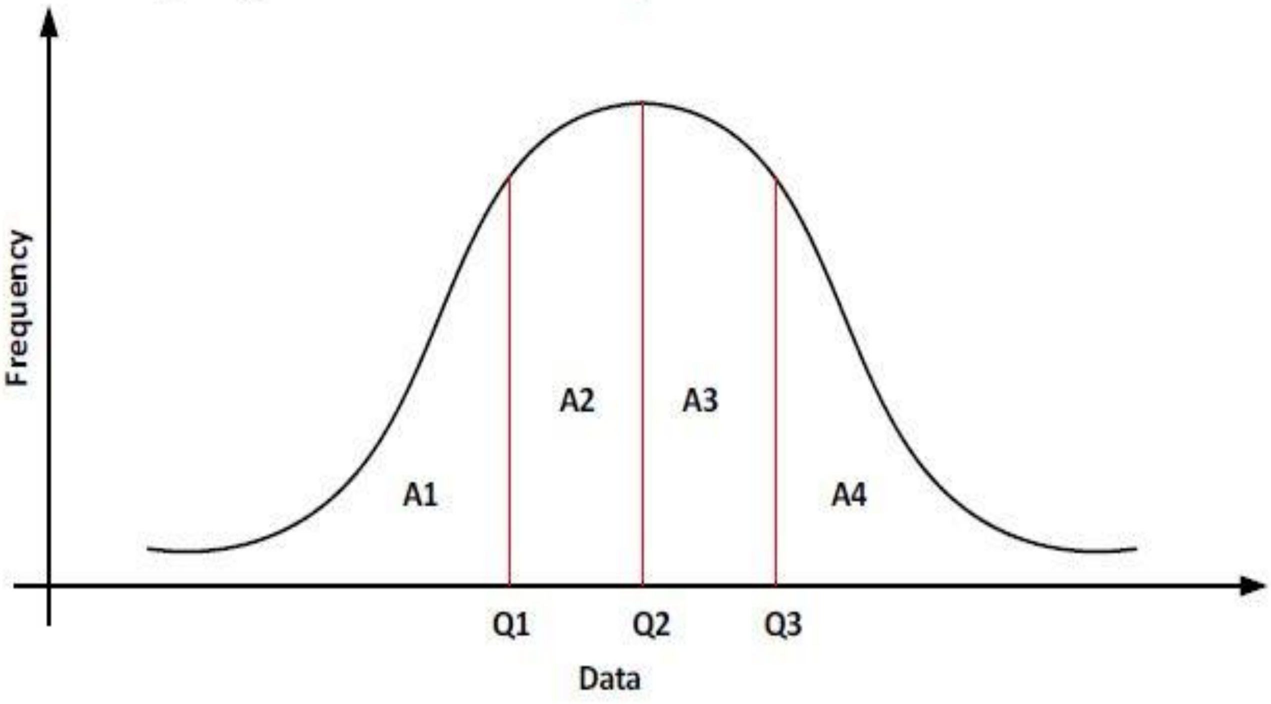
- Just like the median divides the data into half so that 50% of the measurement lies below the median and 50% lies above it, the quartile breaks down the data into quarters so that 25% of the measurements are less than the lower quartile, 50% are less than the median, and 75% are less than the upper quartile.

* Each quartile contains 25% of the total observations. Generally, the data is arranged from smallest to largest:

* **First quartile**: the lowest 25% of numbers
* **Second quartile**: between 25.1% and 50% (up to the median)
* **Third quartile**: 50.1% to 75% (above the median)
* **Fourth quartile:** the highest 25% of numbers

Median and Quartiles





Q1 = first quartile
Q2 = second quartile
Q3 = third quartile
Where:
A1 = A2 = A3 = A4

Quartile for ungrouped data

If there are n-values arranged in ascending order, then Q_1, Q_2 and Q_3 are computed as

$$Q_1 = \text{1st Quartile} = \left(\frac{n+1}{4}\right)\text{th value}$$

$$Q_2 = \text{2nd Quartile} = 2\left(\frac{n+1}{4}\right)\text{th value}$$

$$Q_3 = \text{3rd Quartile} = 3\left(\frac{n+1}{4}\right)\text{th value}$$

Example

Example 1

Find Q_1, Q_2 and Q_3 for the following data.

2, 3, 3, 9, 6, 6, 12, 11, 8, 2, 3, 5, 7, 5, 4, 4, 5, 12, 9

Solution :

First, arrange the data in ascending order:

2, 2, 3, 3, 3, 4, 4, 5, 5, 5, 6, 6, 7, 8, 9, 9, 11, 12, 12

5th 10th 15th

Here $n = 19$, then $Q_1 = \left(\frac{n+1}{4}\right)$th value $= \left(\frac{19+1}{4}\right) =$ 5th value $= 3$

$Q_1 = 3$

$Q_2 = 2\left(\frac{n+1}{4}\right)$th value $= 2\left(\frac{19+1}{4}\right) =$ 10th value $= 5$

$Q_2 = 5$

$Q_3 = 3\left(\frac{n+1}{4}\right)$th value $= 3\left(\frac{19+1}{4}\right) =$ 15th value $= 9$

$Q_3 = 9$

Example 2

Example:

Find the median, lower quartile and upper quartile of the following numbers.

12, 5, 22, 30, 7, 36, 14, 42, 15, 53, 25

Solution:

First, arrange the data in ascending order:

5, 7, 12, 14, 15, 22, 25, 30, 36, 42, 53

lower quartile median upper quartile

Median (middle value) = 22

Lower quartile (middle value of the lower half) = 12

Upper quartile (middle value of the upper half) = 36

If there is an even number of data items, then we need to get the average of the middle numbers.

Example 3

- Find the median, lower quartile, upper quartile, interquartile range and range of the following numbers.
  12, 5, 22, 30, 7, 36, 14, 42, 15, 53, 25, 65

First, arrange the data in ascending order:

5, 7, 12, 14, 15, 22, 25, 30, 36, 42, 53, 65

lower quartile or first quartile

median or second quartile

upper quartile or third quartile

(n+1)/4=3.25

Lower quartile or first quartile $= \frac{12+14}{2} = 13$

Median or second quartile $= \frac{22+25}{2} = 23.5$

Upper quartile or third quartile $= \frac{36+42}{2} = 39$

Interquartile range = Upper quartile – lower quartile
= 39 – 13 = 26

Range = largest value – smallest value
= 65 – 5 = 60

Question

* Find the median, lower quartile and upper quartile of the following numbers.

59, 60, 65, 65, 68, 69, 70, 72, 75, 75, 76, 77, 81, 82, 84, 87, 90, 95, 98

Quartile for grouped data

In case of a continuous (or a grouped) frequency distribution, the Quartiles are computed as

$$Q_1 = l + \frac{h}{f}\left(\frac{\sum f}{4} - C.f.\right) \qquad Q_2 = l + \frac{h}{f}\left(\frac{2\sum f}{4} - C.f.\right)$$

$$Q_3 = l + \frac{h}{f}\left(\frac{3\sum f}{4} - C.f.\right)$$

From the following grouped frequency distribution, Calculate Q_1 and Q_3

Wages in Rs.	150 - 170	170 - 190	190 - 210	210 - 230	230 - 250
No. of workers	30	50	80	30	10

Solution :

C.B.	f	C.f.
150 - 170	30	30
170 - 190	50	80
190 - 210	80	160
210 - 230	30	190
230 - 250	10	200
Total	200	

50th value

150th value

Since $Q_1 = \left(\frac{\sum f}{4}\right)$th value $= \left(\frac{200}{4}\right)$th value = 50th value.

Therefore, class of Q_1 is (170 - 190)

$$Q_1 = l + \frac{h}{f}\left(\frac{\sum f}{4} - C.f.\right)$$

$$Q_1 = 170 + \frac{20}{50}(50 - 30) = 178$$

$Q_1 = 178$

Since $Q_3 = 3\left(\frac{\sum f}{4}\right)$th value $= 3\left(\frac{200}{4}\right)$th value = 150th value

Therefore, class of Q_3 is (190 – 210)

Q3=207.5

Example:

find the the quartiles Q1, Q2, and Q3 of the following data .

Class Interval	Frequency (fi)
50 - 69	3
70 - 89	7
90 - 109	4
110 – 129	4
130 – 149	9

Class Interval	Frequency (fi)	Cumulative frequency	Real interval	
50 - 69	3	3	49.5 – 69.5	
70 - 89	7	10	69.5 – 89.5	$Q1$
90 - 109	4	14	89.5 – 109.5	$Q2$
110 – 129	4	18	109.5 – 129.5	
130 – 149	9	27	129.5 – 149.5	$Q2$

Q1=80.2

Q2=107

Q3=134.5

Question

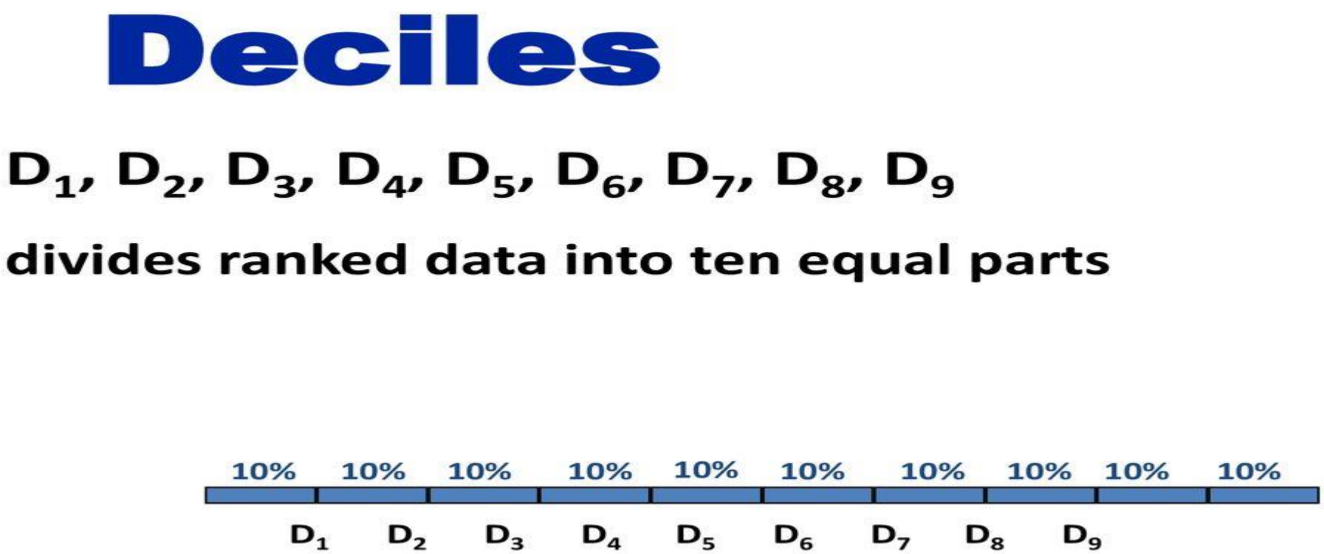
* Find Q1 for following continues group

Wages	1-10	10- 20	20- 30	30-40	40-50	
frequency	22	38	46	35	19	N=160
C F	22	60	106	141	160	

* 14.74

DECILES

- **Deciles divides a data into 10 equal parts. For any series 0f data set , there are 9 deciles denoted by D1,D2…D9. These are called as first decile, second decile so on**



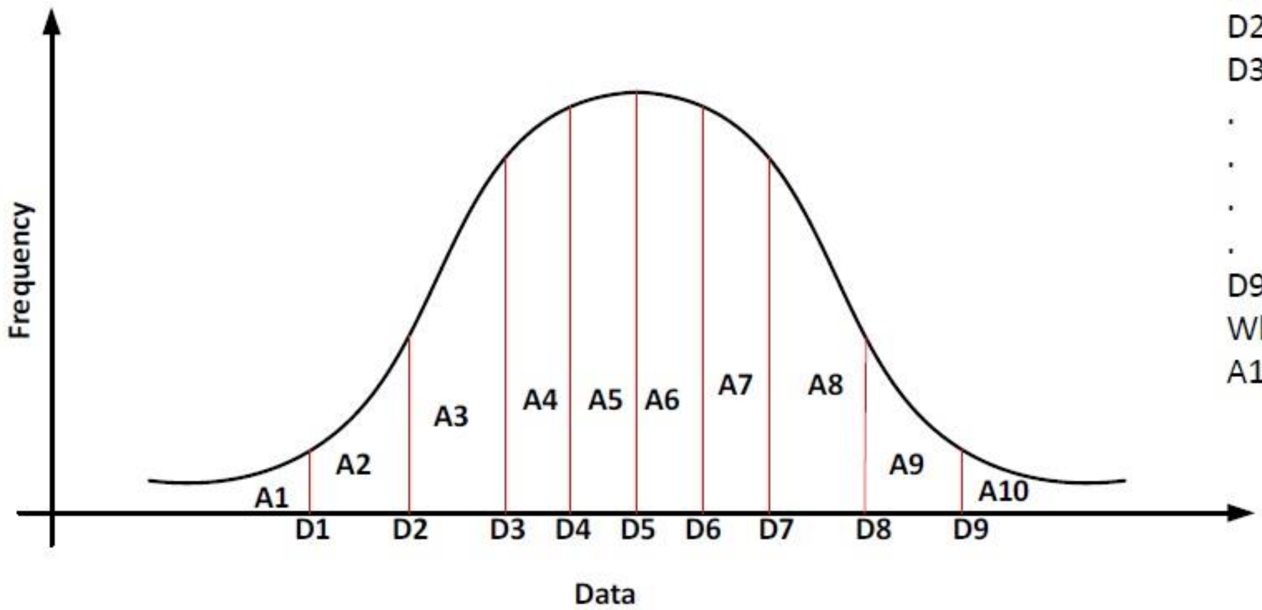
DECILE

* A decile is a quantitative method of splitting up a set of ranked data into 10 equally large subsections.

* A decile rank arranges the data in order from lowest to highest and is done on a scale of one to 10 where each successive number corresponds to an increase of 10 percentage points.

* This type of data ranking is performed as part of many academic and statistical studies in the finance and economics fields.

DECILE



D1 = first decile

D2 = second decile

D3 = third decile

.

.

.

.

D9 = ninthe decile

Where:

A1 = A2 = A3 = A4= . . . .=A10

Decile for ungrouped data

If there are n-values arranged in ascending order, then $D_1, D_2, D_3, \ldots\ldots D_9$ are computed as

$$D_1 = \text{1st Decile} = \left(\frac{n+1}{10}\right)\text{th value}$$

$$D_2 = \text{2nd Decile} = 2\left(\frac{n+1}{10}\right)\text{th value}$$

$$D_3 = \text{3rd Decile} = 3\left(\frac{n+1}{10}\right)\text{th value}$$

. . . . . .

. . . . . .

$$D_9 = \text{9th Decile} = 9\left(\frac{n+1}{10}\right)\text{th value}$$

Example

Example 1

Find D_4 and D_6 from the following weights in kg.

19, 27, 24, 39, 57, 44, 56, 50, 59, 67, 62, 42, 47, 60, 26, 34, 57, 51, 59, 45

Solution :

First we array the data. i.e. 19, 24, 26, 27, 34, 39, 44, 44, 45, 47, 50, 51, 56, 57, 57, 59, 59, 60, 62, 67. Here n = 20

$$D_4 = 4\left(\frac{n+1}{10}\right)\text{th value}$$

$$D_4 = 4\left(\frac{20+1}{10}\right)\text{th value} = 8.4\text{th value} = 8\text{th value} + 0.4[\,9\text{th value} - 8\text{th value}]$$

$$= 44 + 0.4\,[45 - 44] = 44 + 0.4 = 44.4 \text{ kg.}$$

$$D_4 = 44.4 \text{ kg}$$

$$D_6 = 6\left(\frac{n+1}{10}\right)\text{th value} = 6\left(\frac{20+1}{10}\right)\text{th value} = 12.6\text{ th value}$$

$$D_6 = 12\text{th value} + 0.6\,[\,13\text{th value} - 12\text{th value}] = 51 + 0.6\,[\,56 - 51] = 54 \text{ kg.}$$

$$D_6 = 54\text{Kg.}$$

Example 2

Decile for grouped data

In case of a Grouped Frequency distribution, the Deciles are computed as

$$D_1 = l + \frac{h}{f}\left(\frac{\sum f}{10} - C.f.\right)$$

$$D_2 = l + \frac{h}{f}\left(\frac{2\sum f}{10} - C.f.\right)$$

$$D_3 = l + \frac{h}{f}\left(\frac{3\sum f}{10} - C.f.\right)$$

. . . .

. . . .

$$D_9 = l + \frac{h}{f}\left(\frac{9\sum f}{10} - C.f.\right)$$

Example

Calculate D_2 and D_3 from the following data:

x	0 - 5	5 - 10	10 - 15	15 - 20	20 - 25
f	7	18	25	30	20

Solution :

C.B.	f	C.f.
0 - 5	7	7
5 - 10	18	25
10 - 15	25	50
15 - 20	30	80
20 - 25	20	100
Total	100	

20th value

50th value

EXAMPLE CONTINUE..

Since $D_2 = 2\left(\frac{\sum f}{10}\right)$th value $= 2\left(\frac{100}{10}\right)$th value $=$ 20th value.

Therefore, D_2 lies in the class (5 – 10), then

$$D_2 = l + \frac{h}{f}\left(\frac{2\sum f}{10} - C.f.\right) = 5 + \frac{5}{18}(20-7) = 8.6$$

$D_2 = 8.6$

Since $D_3 = 3\left(\frac{\sum f}{10}\right)$th value $= 3\left(\frac{100}{10}\right)$th value $=$ 30th value .

Therefore, D_3 lies in the class (10 – 15), then

$$D_3 = l + \frac{h}{f}\left(\frac{3\sum f}{10} - C.f.\right) = 10 + \frac{5}{25}(30-25) = 11.0$$

$D_3 = 11.0$

Example:

find the desiles D1, D5, and D9 of the following data .

Columns Load	Frequency (fi)
50 - 69	3
70 - 89	7
90 - 109	4
110 – 129	4
130 – 149	9

Class Interval	Frequency (fi)	Cumulative frequency	Real interval	
50 - 69	3	3	49.5 – 69.5	$D1$
70 - 89	7	10	69.5 – 89.5	
90 - 109	4	14	89.5 – 109.5	$D2$
110 – 129	4	18	109.5 – 129.5	
130 – 149	9	27	129.5 – 149.5	$D9$

D1=69.5
D5=107
D9=143.2

PERCENTILES

- • Percentiles are used to understand and interpret data. They indicate the values below which a certain percentage of the data in a data set is found.

- • Percentiles can be calculated using the formula n = (P/100) x N, where P = percentile, N = number of values in a data set (sorted from smallest to largest), and n = ordinal rank of a given value.

- • Percentiles are frequently used to understand test scores and biometric measurements.

What Percentile means?

* Percentiles should not be confused with percentages. The latter is used to express fractions of a whole, while percentiles are the values below which a certain percentage of the data in a data set is found. In practical terms, there is a significant difference between the two. For example, a student taking a difficult exam might earn a score of 75 percent. This means that he correctly answered every three out of four questions. A student who scores in the 75th percentile, however, has obtained a different result. This percentile means that the student earned a higher score than 75 percent of the other students who took the exam. In other words, the percentage score reflects how well the student did on the exam itself; the percentile score reflects how well he did in comparison to other students.

Percentile for Ungrouped data

If there are n-values arranged in ascending order, then $P_1, P_2, P_3,$ ………… P_{99}. are computed as

$$P_1 = \text{1st percentile} = \left(\frac{n+1}{100}\right) \text{th value}$$

$$P_2 = \text{2nd percentile} = 2\left(\frac{n+1}{100}\right) \text{th value}$$

$$P_3 = \text{3rd percentile} = 3\left(\frac{n+1}{100}\right) \text{th value}$$

" " " " " "

" " " " " "

$$P_{99} = \text{99th percentile} = 99\left(\frac{n+1}{100}\right) \text{th value}$$

Find P25, p59 and p80

First we arrange the data i.e.

10, 10, 10, 11, 12, 13, 13, 14, 14, 14, 15, 15, 16, 16, 17, 20

Here $n = 16$

then $P_{25} = 25\left(\frac{n+1}{100}\right)$th value

$$P_{25} = 25\left(\frac{16+1}{100}\right)\text{th value} = 4.25\text{th value}$$

$$= 4\text{th value} + 0.25\,[\,5\text{th value} - 4\text{th value}]$$

$$= 11 + 0.25\,[\,12 - 11] = 11.25$$

$$P_{59} = 59\left(\frac{n+1}{100}\right)\text{th value}$$

$$P_{59} = 59\left(\frac{16+1}{100}\right)\text{th value} = 10.03\text{th value}$$

$$= 10\text{th value} + 0.03\,[\,11\text{th value} - 10\text{th value}]$$

$$= 14 + 0.03\,[\,15 - 14] = 14.03$$

Now $P_{80} = 80\left(\frac{n+1}{100}\right)$th value

$$P_{80} = 80\left(\frac{16+1}{100}\right)\text{th value} = 13.6\text{th value}$$

$$= 13\text{th value} + 0.6\,[14\text{th value} - 13\text{th value}]$$

$$= 16 + 0.6\,[\,16 - 16] = 16$$

Percentile for Grouped Data

$P_{99} = 99\left(\frac{\quad}{100}\right)$

In case of a continuous (or a grouped) frequency distribution, the percentiles are computed as

$$P_1 = l + \frac{h}{f}\left(\frac{\sum f}{100} - C.f.\right)$$

$$P_2 = l + \frac{h}{f}\left(\frac{2\sum f}{100} - C.f.\right)$$

$$P_3 = l + \frac{h}{f}\left(\frac{3\sum f}{100} - C.f.\right)$$

. . . .

. . . .

$$P_{99} = l + \frac{h}{f}\left(\frac{99\sum f}{100} - C.f.\right)$$

Note : It is to be noted that; Median $= Q_2 = D_5 = P_{50}$

Find P20 and P60

Groups	f	$C.f.$
2 - 12	2	2
12 - 22	5	7
22 - 32	8	15
32 - 42	12	27
42 - 52	15	42
52 - 62	20	62
62 - 72	16	78
72 - 82	14	92
82 - 92	10	102
92 - 102	8	110
Total	110	

22nd value

66th value

Since, $P_{20} = 20\left(\frac{\sum f}{100}\right)$th value $= 20\left(\frac{110}{100}\right)$th value $=$ 22nd value

Therefore, P_{20} lies in the group (32 – 40).

Hence; $P_{20} = l + \frac{h}{f}\left(\frac{20\sum f}{100} - C.f.\right)$

$$P20 = 32 + \frac{10}{12}(22-15) = 37.83$$

$$P20 = 37.83$$

Since $P_{60} = 60\left(\frac{\sum f}{100}\right)$th value $= 60\left(\frac{110}{100}\right)$th value $=$ 66th value.

Therefore, P_{60} lies in the group (62 – 72)

Hence; $P_{60} = l + \frac{h}{f}\left(\frac{60\sum f}{100} - C.f.\right)$

10

Example:

find the percentiles P8, P50, and P85 of the following data .

Columns Load	Frequency (fi)
50 - 69	3
70 - 89	7
90 - 109	4
110 – 129	4
130 – 149	9

Solution: 1) find the cumulative frequency and the summation of frequencies and real interval limit.

Columns Load	Frequency (fi)	Cumulative frequency	Real interval	
50 - 69	3	3	49.5 – 69.5	$P8$
70 - 89	7	10	69.5 – 89.5	
90 - 109	4	14	89.5 – 109.5	
110 – 129	4	18	109.5 – 129.5	
130 – 149	9	27	129.5 – 149.5	

P8=69.5
P50=107
P85=140.5

Home work

2 For the following data, find the lower Quartile Q_1 and the upper Quartile Q_3

53 62 79 50 48 80 55 59 63 74 73

3 Calculate Median and Quartiles for the data given below

Marks	30-39	40-49	50-59	60-69	70-79	80-89	90-99
Frequency	2	3	11	20	32	25	7

4 Calculate Q_1, D_4, and P_{95} from the following distribution

C.I.	0-9	10-19	20-29	30-39	40-49	50-59
f	7	13	22	11	6	5

5 The following frequency table gives the height (in inches) of 100 students in a college.

C.I.	60-62	62-64	64-66	66-68	68-70	70-72	Total
f	5	18	42	20	8	7	100

Calculate: (a) Median (b) Q_3 (c) D_7 (d) P_{69}