Measures of Central Tendency

❖ Measures of Central tendency

The measures, which usually reflect the complete data, set and falls in the center of the array is known as measures of central tendency. It may also be called a center or location of the distribution. Such measures are

- Mean
- Median
- Mode

Mean is also there types

- Arithmetic mean
- Geometric mean and
- Harmonic mean

❖ Requisites of a good measures of central tendency

The requirements of ideal measure of central tendency are:

- It should be rigidly defined.
- It should be easy to understand and calculate.
- It should be based on all the observations.
- It should be suitable for further mathematical treatment.
- It should be affected as little as possible by fluctuations of sampling.
- It should not be affected much by extreme observations.

❖ Arithmetic Mean

The arithmetic mean, or the mean, of a set of data is the measure of center found by adding the data values and dividing the total by the number of data values.

Let $x_1, x_2, x_3, \cdots\cdots\cdots, x_n$ be a set of n measurements of a variable x. Then arithmetic mean (AM) is defined as

$$AM; \bar{x} = \frac{x_1 + x_2 + x_3 + \cdots\cdots\cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

For continuous frequency distribution, if $x_1, x_2, x_3, \cdots\cdots\cdots, x_k$ $(i = 1,2,3, \ldots\ldots, k)$ be the mid value of i^{th} class and $f_1, f_2, f_3, \cdots\cdots\cdots, f_k$ be their corresponding frequency. Then AM is computed as

$$AM; \bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{N}\text{ ; Where } N = f_1 + f_2 + f_3 + \cdots\cdots\cdots + f_k.$$

❖ **Example #1:** The marks of 5 students in a class are shown here: 40, 55, 75, 65, 80. Find mean.

Solution: We know that

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$= \frac{40+55+75+65+80}{5}$$

$$= \frac{315}{5} = 63$$

Therefore, the average marks of the students is 63.

Example #2: The following is a distribution of the final examination scores, which 200 students obtained in a three –week course in Statistics.

Scores	0-20	20-40	40-60	60-80	80-100	Total
Number of students	24	55	76	32	13	200

Find the arithmetic mean / average marks of the above distribution.

Solution:

Scores	Number of students (f_i)	Mid value (X_i)	f_iX_i
0-20	24	10	240
20-40	55	30	1650
40-60	76	50	3800
60-80	32	70	2240
80-100	13	90	1170
Total	200		9100

We know that

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i X_i}{N}$$

$$= \frac{9100}{200} = 45.5$$

Therefore, the average marks of the students in Statistics is 45.5.

❖ **Properties of Arithmetic Mean**

Some important properties of the arithmetic mean are as follows:

- The sum of deviations of the items from their arithmetic mean is always zero, i.e. $\sum(X_i - \bar{X}) = 0$.
- The sum of the squared deviations of the items from Arithmetic Mean (A.M) is minimum, which is less than the sum of the squared deviations of the items from any other values. i.e. $\sum(X_i - \bar{X})^2 < \sum(X_i - A)^2$, $Where\ \bar{X} \neq A$.
- Arithmetic mean is depends on change of origin and scale of measurement.

❖ **Merits of Arithmetic Mean**

- The arithmetic mean is simple to understand and easy to calculate.
- It is based on all observations in a series.
- A.M is rigidly defined.
- It has the capability of further algebraic treatment.

❖ Demerits of Arithmetic Mean

- It is changed by extreme items such as very small and very large items.
- In some cases, A.M. does not represent the original item. For example, average patients admitted to a hospital are 10.7 per day.
- The arithmetic mean is not suitable in extremely asymmetrical distributions.
- In case of open ended class intervals, it cannot be calculated without assuming the size of the open end classes.

❖ Geometric Mean (GM)

Geometric mean is the nth root of the product of the data values, where there are n of these. This measure is valid only for data that are measured absolutely on a strictly positive scale.

Let $x_1, x_2, x_3, \cdots\cdots\cdots, x_n$ be a set of n measurements of a variable x. Then geometric mean (GM) is defined as

$$GM = (x_1.x_2.x_3 \cdots\cdots\cdots x_n)^{\frac{1}{n}}$$

For continuous frequency distribution, if $x_1, x_2, x_3, \cdots\cdots\cdots, x_k$ $(i = 1,2,3, \ldots .., k)$be the mid value of i^{th} class and $f_1, f_2, f_3, \cdots\cdots\cdots, f_k$ be their corresponding frequency. Then GM is computed as

$$GM = ({x_1}^{f_1}.{x_2}^{f_2}.{x_3}^{f_3} \ldots\ldots\ldots {x_k}^{f_k})^{\frac{1}{N}} \text{ ; Where } N = f_1 + f_2 + f_3 + \cdots\cdots\cdots + f_k$$

$$\text{Or } GM = Antilog\left(\frac{\sum_{i=1}^{k} f_i log x_i}{N}\right)$$

❖ **Example #3:** Find the G.M of the values 10, 25, 5, and 30.

Solution:
We know that

$$GM = (x_1.x_2.x_3 \cdots\cdots\cdots x_n)^{\frac{1}{n}}$$
$$= (10 \cdot 25 \cdot 5 \cdot 30)^{\frac{1}{4}}$$
$$= (37500)^{\frac{1}{4}} = 13.915$$

Therefore, the geometric mean is 13.915.

Example #4: The following is a distribution of the final examination scores, which 200 students obtained in a three –week course in Statistics.

Scores	0-20	20-40	40-60	60-80	80-100	Total
Number of students	24	55	76	32	13	200

Find the geometric mean of the above distribution.

Solution:

Scores	Number of students (f_i)	Mid value (X_i)	$f_i log X_i$
0-20	24	10	55.26
20-40	55	30	187.07
40-60	76	50	297.31
60-80	32	70	135.95
80-100	13	90	58.50
Total	200		734.09

We know that the geometric mean for the grouped data is

$$GM = Antilog\left(\frac{\sum_{i=1}^{k} f_i log X_i}{N}\right)$$

$$= Antilog\left(\frac{734.09}{200}\right)$$

$$= 39.27$$

Therefore, the geometric mean of the distribution is 39.27.

❖ Merits of Geometric Mean

1. It is rigidly defined and its value is a precise figure.
2. It is based on all observations.
3. It is not affected by the extreme items in the series.
4. It is capable of further algebraic treatment.

❖ Demerits of Geometric Mean

1. It is difficult to compute.
2. It is not easy to understand.
3. If there are negative values in the series, it can not be computed.

❖ Uses of Geometric Mean

- To find the rate of population growth and the rate of interest.
- In the construction of index numbers.

❖ Harmonic Mean (HM)

Harmonic mean is the reciprocal of the arithmetic mean of the reciprocals of the data values. This measure too is valid only for data that are measured absolutely on a strictly positive scale.
Let $x_1, x_2, x_3, \cdots\cdots\cdots, x_n$ be a set of n measurements of a variable x. Then harmonic mean (HM) is defined as

$$HM = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

For continuous frequency distribution, if $x_1, x_2, x_3, \cdots\cdots\cdots, x_k$ $(i = 1,2,3, \ldots\ldots, k)$be the mid value of i^{th} class and $f_1, f_2, f_3, \cdots\cdots\cdots, f_k$ be their corresponding frequency. Then HM is computed as

$$HM = \frac{N}{\sum_{i=1}^{k} \frac{f_i}{x_i}}; \text{ Where } N = f_1 + f_2 + f_3 + \cdots\cdots\cdots + f_k.$$

❖ Example #5:

A train moves first 50 km at a speed of 60 km/ hour, second 50 km at a speed of 75 km/ hour, third 50 km at a speed of 65 km/ hour and fourth 50 km at a speed of 80 km/ hour. What is the average speed of train throughout the journey?

Solution:

Let X denote the speed of the train.

Here given that, $n = 4, X_1 = 60, X_2 = 75, X_3 = 65 \; and \; X_4 = 80.$

We know that

$$HM = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}$$

$$= \frac{4}{\frac{1}{60}+\frac{1}{75}+\frac{1}{65}+\frac{1}{80}} = 69.10$$

Therefore, the average speed of train throughout the journey is 69.10.

Example #6: The following is a distribution of the final examination scores, which 200 students obtained in a three –week course in Statistics.

Scores	0-20	20-40	40-60	60-80	80-100	Total
Number of students	24	55	76	32	13	200

Find the harmonic mean of the above distribution.

Solution:

Scores	**Number of students (f_i)**	**Mid value (X_i)**	**f_i/X_i**
0-20	24	10	2.40
20-40	55	30	1.83
40-60	76	50	1.52
60-80	32	70	0.46
80-100	13	90	0.14
Total	200		6.35

We know that the harmonic mean for the grouped data is

$$HM = \frac{N}{\sum_{i=1}^{k} \frac{f_i}{X_i}}$$

$$= \frac{200}{6.35} = 31.47$$

Therefore, the harmonic mean of the distribution is 31.47.

❖ **Merits of Harmonic Mean**

1. Like AM and GM, it is also based on all observations.
2. It is most appropriate average under conditions of wide variations among the items of a series since it gives larger weight to smaller items.
3. It is capable of further algebraic treatment.
4. It is extremely useful while averaging certain types of rates and ratios.

❖ **Demerits of Harmonic Mean**

1. It is difficult to understand and to compute.
2. It cannot be computed when one of the values is 0 or negative.
3. It is necessary to know all the items of a series before it can be calculated.

- **Uses of Harmonic Mean**

If there are two measurements taken together to measure a variable, HM can be used. For example, tonne mileage, speed per hour. In the above example tonne mileage, tonne is one measurement and mileage is another measurement. HM is used to calculate average speed.

- **Theorem #1:** Show that the sum of deviations of the items from their arithmetic mean is always zero, i.e. $\sum(X_i - \bar{X}) = 0$.

Solution:

We have, $\sum(X_i - \bar{X}) = \sum X_i - \sum \bar{X}$

$$= \sum X_i - n\bar{X}$$

$$= n\bar{X} - n\bar{X} = 0 \text{ (Showed)} \qquad \left[\text{Since } \bar{X} = \frac{\sum X_i}{n} \Rightarrow \sum X_i = n\bar{X}\right]$$

- **Theorem #2:** Show that arithmetic mean is depends on change of origin and scale of measurement.

Solution:

Let $x_1, x_2, x_3, \cdots\cdots\cdots, x_n$ be a set of n measurements of a variable x. Then arithmetic mean (AM) is

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Let us consider a new variable $u_i = \frac{x_i - a}{c}$, $where\ (i = 1, 2, 3, \dots\dots\dots, n)$ and a and c are arbitrary constant denotes the origin and scale respectively.

$$x_i = a + cu_i$$

$$\bar{x} = \frac{\sum_{i=1}^{n} a + cu_i}{n}$$

$$= \frac{\sum_{i=1}^{n} a + c\sum_{i=1}^{n} u_i}{n}$$

$$= \frac{na}{n} + c\frac{\sum_{i=1}^{n} u_i}{n}$$

$$= a + c\bar{u} \qquad \left[\text{Since}\bar{u} = \frac{\sum_{i=1}^{n} u_i}{n}\right]$$

Therefore, arithmetic mean is depends on change of origin and scale of measurement. (Showed)

- **Relationship among Arithmetic mean (AM), Geometric mean (GM) and Harmonic mean (HM)**

  - $AM \geq GM \geq HM$
  - $AM \times HM = (GM)^2$

Theorem #3: For two non-negative observations, show that $AM \geq GM \geq HM$.
Proof: Let x_1 and x_2 be two non-negative values of a variable x.

Therefore,

$$AM = \frac{x_1 + x_2}{2}$$

$$GM = \sqrt{x_1 x_2}$$

$$HM = \frac{2}{\frac{1}{x_1}+\frac{1}{x_2}} = \frac{2}{\frac{x_1+x_2}{x_1 x_2}} = \frac{2x_1x_2}{x_1+x_2}$$

$$AM - GM = \frac{x_1 + x_2}{2} - \sqrt{x_1 x_2}$$

$$= \frac{x_1 + x_2 - 2\sqrt{x_1 x_2}}{2}$$

$$= \frac{(\sqrt{x_1})^2 + (\sqrt{x_2})^2 - 2\sqrt{x_1 x_2}}{2}$$

$$= \frac{(\sqrt{x_1} - \sqrt{x_2})^2}{2}$$

= +ve term (if $x_1 \neq x_2$)

$$AM - GM > 0$$

$$AM > GM \cdots\cdots (i)$$

If $x_1 = x_2$ then

$$AM - GM = 0$$

$$AM = GM \cdots\cdots\cdots (ii)$$

From equation (i) and (ii) we can write

$$AM \geq GM \cdots\cdots\cdots (iii)$$

$$GM - HM = \sqrt{x_1 x_2} - \frac{2x_1x_2}{x_1 + x_2}$$

$$= \frac{\sqrt{x_1 x_2}(x_1 + x_2) - 2x_1x_2}{x_1 + x_2}$$

$$= \frac{\sqrt{x_1 x_2}}{x_1 + x_2}\left(x_1 + x_2 - 2\sqrt{x_1 x_2}\right)$$

$$= \frac{\sqrt{x_1 x_2}}{x_1 + x_2}(\sqrt{x_1} - \sqrt{x_2})^2$$

$$= +\text{ve term (if } x_1 \neq x_2)$$

$$GM - HM > 0$$

$$GM > HM \cdots\cdots\cdots (iv)$$

If $x_1 = x_2$ then

$$GM - HM = 0$$

$$GM = HM \cdots\cdots\cdots (v)$$

From equation (iv) and (v) we can write

$$GM \geq HM \cdots\cdots\cdots (vi)$$

From equation (iii) and (vi) we can write

$$AM \geq GM \geq HM \quad \text{(Showed)}$$

Theorem #4: A series of values with a common ratio r as follows $a, ar, ar^2, ar^3, \ldots\ldots\ldots, ar^{n-1}$. Find arithmetic mean, geometric mean and harmonic mean of the series and show that $AM \times HM = (GM)^2$.

Solution:

Given that, the geometric series with common ratio r is as follows

$$a, ar, ar^2, ar^3, \ldots\ldots\ldots, ar^{n-1}$$

We have that

$$AM = \frac{a + ar + ar^2 + ar^3 + \cdots\cdots\cdots + ar^{n-1}}{n}$$

$$= \frac{a}{n}(1 + r + r^2 + r^3 + \cdots\cdots\cdots + r^{n-1})$$

$$= \frac{a}{n} \cdot \frac{1-r^n}{1-r} \text{ since } r < 1$$

$$\text{GM} = (a.ar.ar^2.ar^3 \ldots \cdots\cdots . ar^{n-1})^{\frac{1}{n}}$$

$$= (a^n . r^{1+2+3+\cdots\cdots+n-1})^{\frac{1}{n}}$$

$$= a.(r^{\frac{(n-1)(n-1+1)}{2}})^{\frac{1}{n}}$$

$$= ar^{\frac{n-1}{2}}$$

$$HM = \frac{n}{\frac{1}{a} + \frac{1}{ar} + \frac{1}{ar^2} + \frac{1}{ar^3} + \cdots\cdots + \frac{1}{ar^{n-1}}}$$

$$= \frac{n}{\frac{1}{a}(1 + \frac{1}{r} + \frac{1}{r^2} + \frac{1}{r^3} + \cdots\cdots + \frac{1}{r^{n-1}})}$$

$$= \frac{an}{\frac{(\frac{1}{r})^n - 1}{\frac{1}{r} - 1}}$$

$$=\frac{an}{\frac{1-r^n}{r^n}\cdot\frac{r}{1-r}}$$

$$=\frac{an(1-r)r^{n-1}}{1-r^n}$$

$$AM \times HM = \frac{a}{n}\cdot\frac{1-r^n}{1-r}\times\frac{an(1-r)r^{n-1}}{1-r^n}$$
$$= a^2r^{n-1}$$
$$= (ar^{\frac{n-1}{2}})^2$$
$$= (GM)^2$$

Therefore $AM \times HM = (GM)^2$ (Showed)

- **Median**

Median is the middle value of an array or a series, which divides the array into two equal parts half of the observations are above it, and half of the observations are below it.

- **Computation of Median**
- Arrange the series in ascending or descending order.
- If the number of observation (n) is odd, then the formula of median is
  Median = Value of $\left(\frac{n+1}{2}\right) th$ observation
- If the number of observation (n) is even, then the formula of median is
  Median = Value of $\frac{1}{2}\left(\left(\frac{n}{2}\right) th\ observation + \left(\frac{n}{2}+1\right) th\ observation\right)$

- **Example #7:** Find the median of the series: 2, 6, 7, 4, and 11.

Solution:

By arranging the series in ascending order, we have that

2, 4, 6, 7, and 11

Since the number of observation $(n = 5)$ is odd , therefore

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

= Value of $\left(\frac{5+1}{2}\right) th$ observation

= Value of $3rd$ observation = 6

- **Example #8:** Find the median of the series: 2, 6, 7, 4, 9, and 11.

Solution:

By arranging the series in ascending order, we have that

2, 4, 6, 7, 9, and 11

Since the number of observation $(n = 6)$ is even, therefore

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

$$= \frac{1}{2} \text{ Value of } \left(\left(\frac{6}{2}\right) th\ observation + \left(\frac{6}{2}+1\right) th\ observation\right)$$

$$= \frac{1}{2} \text{ Value of } (\ 3rd\ observation + 4th\ observation)$$

$$= \frac{1}{2}\ (6 + 7) = 6.5$$

- **Computation of Median from grouped data**

$$\text{Median} = L + \frac{\frac{N}{2} - F}{f}\ i$$

Where,

Median class = which class have $cf \geq \frac{N}{2}$.

L is the lower limit of median class

N is the total frequency

f is the frequency of median class

F is the cumulative frequency preceding of the median class

i is the class width of median class

- **Example #9:** The following is a distribution of the final examination scores, which 200 students obtained in a three –week course in Statistics.

Scores	0-20	20-40	40-60	60-80	80-100	Total
Number of students	24	55	76	32	13	200

Find the median of the distribution.

Solution:

Scores	**Number of students (f)**	**Cumulative frequency**
0-20	24	24
20-40	55	79
40-60	76	155
60-80	32	187
80-100	13	200
Total	200	

We know that the median for grouped data is

$$\text{Median} = L + \frac{\frac{N}{2} - F}{f}\ i$$

$$= 40 + \frac{\frac{200}{2} - 79}{76} \times 20$$

$$= 40 + 5.53$$

$$= 45.53$$

Where,

Median class = which class have $cf \geq \frac{N}{2} = 40 - 60$

L is the lower limit of median class

N is the total frequency

f is the frequency of median class

F is the cumulative frequency preceding of the median class

i is the class width of median class

Therefore, the median of the distribution is 45.53.

❖ Merits of Median

1. It is simple to understand and easy to calculate.
2. It is not affected by the extreme items in the series.
3. It can be determined graphically.
4. For open-ended classes, median can be calculated.
5. It can be located by inspection, after arranging the data in order of magnitude.

❖ Demerits of Median

1. It does not consider all variables because it is a positional average.
2. The value of median is affected more by sampling fluctuations
3. It is not capable of further algebraic treatment.
4. It cannot be computed precisely when it lies between two items.

❖ Uses of Median

1. Median is used when the exact midpoint of the distribution is needed or the 50% point is wanted.
2. When extreme scores affect the mean at that time median is the best measure of central tendency.
3. Median is used when it is required that certain scores should affect the central tendency, but all that is known about them is that they are above or below the median.
4. Median is used when the classes are open-ended or it is of unequal cell size.

❖ Mode

The value that occurs most often in a data set is called the **mode.**

- A data set that has only one value that occurs with the greatest frequency is said to be **unimodal**.
- If a data set has two values that occur with the same greatest frequency, both values are considered to be the mode and the data set is said to be **bimodal**.
- If a data set has more than two values that occur with the same greatest frequency, each value is used as the mode, and the data set is said to be **multimodal**.
- When no data value occurs more than once, the data set is said to have no mode.
- A data set can have more than one mode or no mode at all.

❖ **Example #10:** Find the mode of the following series

(i) 2, 4, 9, 4, 6, 7, 4
(ii) 2, 4, 9, 4, 7, 2, 5
(iii) 2, 4, 5, 7, 9

Solution:

(i) Arrange the series in ascending order: 2, 4, 4, 4, 6, 7, 9
Since 4 appears three times and all the rest appears only once, so the mode is 4.
(ii) Arrange the series in ascending order: 2, 2, 4, 4, 5, 7, 9
Since 2 and 4 appears twice and all the rest appears only once, so the mode is 2 and 4.
(iii) Since all the observations appears once, so there is no mode.

- **Mode for grouped data**

$$Mode = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} i$$

Where,
L is the lower limit of modal class.
Δ_1 is the difference between the frequencies of modal class and pre-modal class.
Δ_2 is the difference between the frequencies of modal class and post-modal class.
i is the class width of modal class.

- **Example #11:** The following is a distribution of the final examination scores, which 200 students obtained in a three –week course in Statistics.

Scores	0-20	20-40	40-60	60-80	80-100	Total
Number of students	24	55	76	32	13	200

Find the mode of the distribution.

Solution:

Scores	**Number of students (f)**
0-20	24
20-40	55
40-60	76
60-80	32
80-100	13
Total	200

We know that the mode for grouped data is

$$\begin{aligned} Mode &= L + \frac{\Delta_1}{\Delta_1 + \Delta_2} i \\ &= 40 + \frac{21}{21 + 44} \times 20 \\ &= 40 + 6.46 \\ &= 46.46 \end{aligned}$$

Where,
Modal class is 40-60 since most frequency occurs in this class.
L = 40
Δ_1= 76-55=21
Δ_2= 76-32=44
i=20

Therefore, the mode of the distribution is 46.46.

- **Merits of Mode**

1. It is easy to understand and simple to calculate.
2. It is not affected by extremely large or small values.
3. It can be useful for qualitative data.
4. It can be computed in an open-end frequency table.
5. It can be located graphically.

- **Demerits of Mode**

1. It is not well defined.
2. It is not based on all the values.
3. It is stable for large values so it will not be well defined if the data consists of a small number of values.
4. It is not capable of further mathematical treatment.

- **Uses of Mode**

1. When we want a quick and approximate measure of central tendency.
2. Mode is useful for qualitative data.
3. When we want a measure of central tendency, which should be typical value. For example when we want to know the typical dress style of Bangladesh women, i.e. the most popular dress style. Like this the average marks of a class is called modal marks.

- **Quartiles**

The quartiles of a data set divide the data into four equal parts, with one-fourth of the data values in each part. Each data set have three quartiles, such as (i) first quartile (Q_1), (ii) second quartile (Q_2) and (iii) third quartile (Q_3). Where Q_1represents the first 25% observations are below it and 75% observations are above it, Q_2 represents the 50% observations, half of the observations are below it and half are above it i.e. $Median = Q_2$ and Q_3 represents the first 75% observations are below it and last 25% observations are above it of a series.

- **Computation of Quartiles**

- Arrange the series in ascending or descending order.
- If the number of observation (n) is odd, then the formula of median is

  Q_i= Value of $\left(\frac{i(n+1)}{4}\right) th$ observation, where $i = 1, 2, 3.$
- If the number of observation (n) is even, then the formula of median is

  Q_i = Value of $\frac{1}{2}((\frac{in}{4})th\ observation + \left(\frac{in}{4} + 1\right) th\ observation)$, where $i = 1, 2, 3.$

- **Example #12:** Find the quartiles of the series: 2, 6, 7, 9, 5, 4, and 11.

Solution:

By arranging the series in ascending order, we have that

2, 4, 5, 6, 7, 9 and 11

Since the number of observation $(n = 7)$ is odd , therefore

Q_i= Value of $\left(\frac{i(n+1)}{4}\right) th$ observation, where $i = 1, 2, 3.$

Q_1= Value of $\left(\frac{(n+1)}{4}\right) th$ observation
= Value of $\left(\frac{7+1}{4}\right) th$ observation
= Value of $2nd$ observation = 4

Q_2= Value of $\left(\frac{(n+1)}{2}\right) th$ observation
= Value of $\left(\frac{7+1}{2}\right) th$ observation
= Value of $4th$ observation = 6

Q_3= Value of $\left(\frac{3(n+1)}{4}\right) th$ observation
= Value of $\left(\frac{3(7+1)}{4}\right) th$ observation
= Value of $6th$ observation = 9

- **Computation of Quartiles from grouped data**

$$Q_i = L_i + \frac{\frac{iN}{4} - F_i}{f_i} \times c \qquad ; i = 1, 2, 3.$$

Where,

Quartile class = which class have $cf \geq \frac{iN}{4}$.

L_i is the lower limit of ith quartile class

N is the total frequency

f_i is the frequency of ith quartile class

F_i is the cumulative frequency preceding of the ith quartile class

c is the class width of quartile class

- **Example #13:** The following is a distribution of the final examination scores, which 200 students obtained in a three –week course in Statistics.

Scores	0-20	20-40	40-60	60-80	80-100	Total
Number of students	24	55	76	32	13	200

Find the quartiles of the distribution.

Solution:

Scores	Number of students (f)	Cumulative frequency
0-20	24	24
20-40	55	79
40-60	76	155
60-80	32	187
80-100	13	200
Total	200	

We know that the quartiles for grouped data is

$$Q_i = L_i + \frac{\frac{iN}{4} - F_i}{f_i} \times c \qquad ; i = 1, 2, 3.$$

Where,

Quartile class = which class have $cf \geq \frac{iN}{4}$.

L_i is the lower limit of ith quartile class

N is the total frequency

f_i is the frequency of ith quartile class

F_i is the cumulative frequency preceding of the ith quartile class

c is the class width of quartile class

1st quartile:

$$Q_1 = L_1 + \frac{\frac{N}{4} - F_1}{f_1} \times c$$

$$= 20 + \frac{\frac{100}{4} - 24}{55} \times 20$$

$$= 20 + 0.36 = 20.36$$

Where,

1st quartile class is 20-40

$L_1 = 20$

$F_1 = 24$

$f_1 = 55$

c=20

2nd quartile:

$$Q_2 = L_2 + \frac{\frac{N}{2} - F_2}{f_2} \times c$$

$$= 40 + \frac{\frac{200}{2} - 79}{76} \times 20$$

$$= 40 + 5.53$$

$$= 45.53$$

Where,
2nd quartile class is 40-60
$L_2 = 40$
$F_2 = 79$
$f_2 = 76$
c = 20

3rd quartile:

$$Q_3 = L_3 + \frac{\frac{3N}{4} - F_3}{f_3} \times c$$

$$= 40 + \frac{\frac{3 \times 200}{4} - 79}{76} \times 20$$

$$= 40 + 18.68$$

$$= 58.68$$

Where,
3rd quartile class is 40-60
$L_2 = 40$
$F_2 = 79$
$f_2 = 76$
c = 20

- **Deciles**
- **Percentiles**